Planning for Active Living:

How do living environments stimulate active behaviour?

David Grein

Master's Thesis for the Spatial Planning program

Planning, Land and Real Estate Development

Nijmegen School of Management

Radboud University Nijmegen

2022

Planning for Active Living: How do living environments stimulate active behaviour?

Colophon

Project
Master's Thesis

Author
David Grein
S1063913

Supervisor
Dr. Kevin Raaphorst

Second reader
Dr. Henk-jan Kooij

Master Program
Master Spatial Planning
Planning, Land and Real Estate Development

University
Nijmegen School of Management
Radboud University

Keywords
Walkability, Active Living, Built Environment, Active Mobility Behaviour

Word count
32.435

Date
29-08-2022

Abstract
The built environment, a lower demand for physical work and the increase in technological innovations have all led to more sedentary behaviour in recent years, which has resulted in an increase in demand for healthcare services. The built environment in particular has a major influence on the degree of active mobility of individuals in both urban and rural areas.

To examine the walkability of a place, several studies have developed walkability indexes. One of these has also been carried out for the Netherlands and this research continues from there.
The current walkability index is fairly accurate for urban areas. However, it lacks validity for rural areas, partly due to the large difference in geographical classifications between urban and rural areas. This research tries to solve this problem and create a more valid walkability index for rural areas.

In addition, this research will investigate what the Dutch walk culture exactly looks like and which conclusions can be drawn based on the updated walkability index in combination with socio-demographic data. The aim is to provide policy makers and urban planners with better tools to design areas for a more active lifestyle.

Summary
The degree of walkability largely determines the state of health of people in a certain area; the worse the walkability, the greater the chance of a sedentary lifestyle. Previous studies have examined the extent of walkability in certain areas. However, these studies focus mainly on urban areas and to a lesser extent on rural areas. Walkability indexes (WI) that work well in urban areas are used similarly in rural areas even though these areas differ greatly. An example is GECCO, which has drawn up a WI for the whole of the Netherlands based on seven components. This WI shows the state of walkability in urban areas. However, the reliability of this index decreases as the urbanisation of an area decreases.

To examine how walkability is related to health, UMC Nijmegen has started a study that examines the relationship between walkability and health data of inhabitants of Drenthe, Friesland and Groningen. The problem that arises is that health data are privacy sensitive, and surveys may not be linked to respondents. The smallest scale on which research can be done are postal codes. However, in rural areas, postal codes are often much larger than in urban areas, while far fewer people live there, resulting in unreliable results. This research therefore tries to evaluate and remodel GECCO's walkability index to provide policymakers and city planners a more valid and reliable index, especially for rural areas as most of the areas in the northern Netherlands are low urbanised.

In this study, a GIS was used to investigate what other methods could be adopted to improve the WI for rural areas. GECCO used a centroid, the centre of a PC6 area, but in large PC6 areas this results in an unreliable result. In this study, two new approaches are tested, one based on all values in a PC6 area (PC6-approach) and one based on the locations of residential properties in a PC6 area (RP-approach).

It was found that both methods work better than the Centroid-approach. The PC6-approach is user friendly and does not require a lot of computing power. However, this method is not always accurate, especially in large PC6 areas with multiple components or PC6 areas that consist of several parts. The RP-approach is the most accurate for determining a WI-score in rural areas. This method assumes places where people live and only the locations of residential properties are used to determine the score per component. Finally, all the scores per component are added up and this results in a new WI-score.

It can be concluded that the results from the RP-approach can be used in health research and that the results are more accurate and reliable than in the Centroid-approach. A disadvantage of this approach is that it is very time-consuming and computationally expensive. In the future, a similar approach could be used using a dataset other than the BAG. However, drawing up a WI based on locations of residential properties seems to be the most reliable for a WI at PC6 level, based on this research.

Preface

I hereby present my master's thesis for the conclusion of the master's programme in Spatial Planning. This research focuses on walkability in combination with health data that represents my progress during this master programme at Radboud University. During the master programme Spatial Planning I followed the specialisation; Planning, Land and Real Estate Development, which I considered to be a useful next step after my higher education; Geomedia & Design. During my studies, I worked a lot with Geographical Information Systems, and I was keen to use this knowledge in my master's thesis. The subject of walkability in relation to the built environment really appealed to me, because I believe that cities should spare the car for a better and more beautiful quality of life in the city. Studying walkability and combining this issue with GIS and statistics seemed a very interesting challenge. This resulted in a more technical thesis in which many analyses had to be conducted, which corresponded well to my previous education.

During this thesis, the issue was mainly about rural areas, where I was initially mainly interested in urban areas, this research has shown me that both urban and rural areas are equally important for policy makers and planners. Especially when health data is linked to geographic data, something that was still new to me. This combination has given me interesting insights and I believe that many more interesting studies can be conducted in this field.

During the thesis, not everything went well, mainly hardware limitations resulted in time-consuming analyses. Additionally, I had not previously worked extensively with grids in GIS, so this also took some time to fully understand. However, my supervisor Kevin Raaphorst supported me well when I thought I was bogged down and together we came up with new possible analyses that ultimately yielded a good result. In addition to geographical analyses, I also worked in SPSS for the statistical basis of the analyses. I had not worked with SPSS for quite some time, but I received a lot of help from Marcia Spoelder-Merkens as a supervisor on location at the Radboud UMC. Therefore, I would like to thank Kevin Raaphorst and Marcia Spoelder-Merkens for the pleasant cooperation and the help that was offered when necessary during this study.

List of abbreviations

BAG – Basisregistratie Adressen en Gebouwen or Addresses and Buildings Registry

BMI – Body Mass Index

DoU – Degree of Urbanisation

GECCO – Geoscience and health cohort consortium

GIS – Geographic Information System

OAD - Omgevingsadressendichtheid" or Surrounding address density

PC6 – Postal Code with four characters and two numbers

RP – Residential Property

VIF - Variance Inflation Factor

WI – Walkability Index

Table of contents

Summary 4

Preface 5

List of abbreviations 6

Table of contents 7

1. Introduction 9

1.1 Research problem statement 9

1.2 Research aim 9

1.3 Research question 10

1.3.1 Sub questions 10

1.4 Scientific relevance 10

1.5 Societal relevance 11

2. Theoretical framework 12

2.1 Literature review 12

2.2 Walkability 14

2.3 Determinants of walkability in the Netherlands 15

2.4 Built environment 16

2.5 Walkability index 18

2.6 Limitations to the walkability index 18

2.8 Conceptual Framework 20

3. Methodology 21

3.1 Research philosophy 21

3.2 Research strategy and methods 21

3.3 GIS Analysis 23

3.4 SPSS & R-Studio analysis 26

3.5 LifeLines 29

3.6 Data collection 30

3.7 Validity 30

3.8 Reliability 31

4. Research Results 32

4.1 Dutch walking culture 32

4.2 500 metres approaches 37

4.2.1 Centroid to PC6 38

4.2.2 Centroid to RP 40

4.2.3 PC6 to RP 42
4.3 1650-metres approaches 46
4.3.1 Centroid to PC6 46
4.3.2 Centroid to RP 47
4.3.3 PC6 to RP 47
4.4 Walkability, social behaviour and geography..51
4.4.1 Degree of Urbanisation 51
4.4.2 Walking time 52
4.4.3 Age..53
4.4.4 Gender..53
4.4.5 Educational level 54
4.4.6 Health 55
4.5 Analyses per WI component 57
4.5.1 Univariate regression 57
4.5.2 Multivariate regression 58
5. Discussion..60
6. Conclusion 64
7. References..67
8. Appendix 71
8.1 Appendix I Technical GIS document 71
8.2 Appendix II regression walking for Leisure and components..76

1. Introduction

1.1 Research problem statement

A healthy living environment is nowadays taken for granted, and lowering emissions, reducing nitrogen levels and setting environmental standards, have been important issues in city planning for many years. But are these interventions enough to ensure that people can live healthier lives?

The built environment, a lower demand for physical work and the increase in technological innovations have all led to more sedentary behaviour in recent years, which has resulted in an increase in demand for healthcare services (Faskunger, 2013). Even though there is a lot of evidence that an active lifestyle leads to better physical health, reduction of chronic diseases and improved resistance compared to people who live more sedentary lives (Cavill, Kahlmeier, & Racioppi, 2006).

In the new Dutch Environment Act, one of the goals is therefore to achieve and maintain a healthy living environment, so health must be incorporated into the new environment policy (Raad voor de leefomgeving en infrastructuur, 2018). The problem with this statement is that it can be interpreted very broadly, making it far from clear to policymakers how it should be implemented. To better implement this issue, policy makers should focus on health promotion. This means creating a living environment that invites healthy behaviour and healthy choices (Raad voor de leefomgeving en infrastructuur, 2018). In other words, designing the built environment in such a way that it encourages people to behave in a healthier way. In policy terms, this means promoting walkability and cyclability by designing the built environment in the most appropriate way. Wang and Yang (2020) describe walkability as the degree to which the built environment invites people to walk, how it contributes to a healthy lifestyle and how it increases the quality of life for citizens, the same applies to cyclability with a focus on biking.

Several studies have been conducted into the degree of walkability, one of which was conducted by the Geoscience and health cohort consortium (GECCO). They have drawn up a walkability index for the Netherlands based on a number of factors. This index gives a reasonable indication of the degree to which a place in the Netherlands is walkable or not. However, the factors considered, and the results of the research are mainly valid for urban areas. As a result, the index is accurate for urban areas but less reliable for rural areas. This is partly due to the geographical classification which is less detailed in rural areas than in urban areas. In addition, the mobility behaviour of individuals in rural areas is not similar as in urban or metropolitan areas. For example, amenities are further apart and utilities are less likely to be within walking distance, nevertheless, differences between different rural areas can still be seen in this index. These differences lead to health disparities between urban and rural areas. This result in the fact that people who live in rural communities are more likely than, their urban counterparts, to die premature from death causes such as cancer, heart disease, unintentional injury or stroke (CDC, 2019). It is therefore imperative to also provide reliable indexes for rural areas, to reduce these disparities between rural and urban areas. Focussing more on rural areas and counteracting the metropolitan bias is one of the aims of this research.

1.2 Research aim

To determine how mobility-friendly cities and rural areas are, this research focuses on the design of the built environment and how it relates to walkability. Part of this research will focus on how we can influence travel behaviour in the city with the aim of promoting active travel behaviour and how it is experienced by users. For example, by identifying the Dutch walking culture and how this differs from urban and rural areas.

To illustrate how travel behaviour is influenced by the Dutch built environment, this research tries to refine the existing walkability index from GECCO to ensure a more valid index for rural areas, based on geographical and socio-demographic characteristics. By validating this index for all geographical areas instead of just urban areas, the index will be much more interesting and usable for policymakers and city planners all over the country. Furthermore, more attention to rural communities is necessary as these communities are more prone to diseases caused by, for example, sedentary lifestyles.

The goal of this research is therefore to improve the reliability and validity of the existing walkability index from GECOO for all Dutch geographical areas. After improving the index, this research tries to identify and compare different neighbourhoods and different target groups by answering questions like, why are there differences between neighbourhoods in, for example, rural areas, or between rural and urban areas. But also, what differences can be found between different target groups? This will be done with the help of data of LifeLines, a research programme of the UMC Groningen, which includes geographical and socio-demographical data for over 30 years from Dutch inhabitants.

1.3 Research question

To what extent can we evaluate and remodel the existing walkability index from GECCO to provide policymakers and city planners a more valid and reliable index for all Dutch geographical areas?

1.3.1 Sub questions

1. What constitutes Walkability in social and cultural terms and between different degrees of urbanisation in the Dutch context?
2. What geographical unit is most suitable for conducting walkability studies to validate the index for all geographical areas in the Netherlands?
3. Which relations can be found between the walkability scores and individuals active mobility behaviour in different degrees of urbanisation and between different target groups?

1.4 Scientific relevance

Both in the Netherlands and abroad, much research has been done into walkability. In several countries, such as China, India and the United States, researchers have tried to draw up a walkability index. In addition, several studies have been conducted that focus on the relationship between the built environment and an active lifestyle.

Examples of studies that focus on this relationship are Hansen, Umstattd, Lenardson, & Hartley (2015), Haybatollahi, Czepkiewicz, Laatikainen, & Kytta (2015), Engelen, Dhillon, Chau, Hespe, & Bauman (2016). In addition, there have been several studies such as Zuniga-Teran, et al. (2017) and Wang & Yuqi (2019) that have established models and indices regarding walkability. Many of these studies focus on foreign cities but also in the Netherlands, Liao, van den Berg, van Wesemael, & Arentze (2020) have previously attempted to establish a walkability model and index.

The problem with these different indices is that they are never completely accurate. Almost every research shows that the correct indexes and variables are not always used, and that follow-up research is necessary. This is partly because the degree of walkability differs per country; American and Chinese cities, for example, are designed differently from Dutch cities and this also translates into different factors that influence the degree of walkability. Although a walkability index for the Netherlands does exist, this index is not validated for all geographical areas in the country.

Therefore, this research will continue to build on previous studies and the GECCO walkability index, while considering the appropriate variables that are specific to the Netherlands. This research is scientifically relevant, because it contributes to an improved version of the walkability index for the

Netherlands, where a distinction is made between different target groups and different degrees of urbanisation. This has not explicitly been considered in previous studies.

1.5 Societal relevance

Health is an important aspect in today's society, and this can be expressed in many ways. Reducing harmful emissions and combating climate change are health aspects that have been frequently mentioned in recent years. However, less often mentioned is the influence of the built environment on human health even though it has a strong influence on the behaviour and choices of citizens in their daily needs. The frequent use of cars and the filling-in of workplaces are leading to an increasingly sedentary existence. The damage to health caused by this has been known for a long time, and guidelines for healthier and walkable friendlies cities do exist. However, their implementation lags behind.

An important factor in counteracting a sedentary lifestyle is to think about how the built environment is designed. Are people invited to take the car or public transport? How many people walk or cycle to work or in their free time? Contemporary cities easily invite people to take the car, whether for work, shopping or outings.

This research will focus on how walking, cycling or using public transport becomes more inviting than using the car and particularly on how the built environment plays a role in this process. Investigating this issue is very relevant to society as it largely determines the health of citizens. After all, health is one of the most important core values for a happy life. As more and more people want to live in cities, it is important that city residents live healthy lives and that the environment in which they live also invites them to do so.

The walkability index that is attempted to be drawn up based on this research can help city planners and policymakers to improve walkability, and therefore which neighbourhoods are friendlier to health, and which are not. The ultimate goal is to create a walkability index valid for all Dutch geographical areas that helps city planners and policymakers to create walkable friendly cities and neighbourhoods in all degrees of urbanisation.

2. Theoretical framework

In this chapter, the concepts and theories central in this research are presented and explained. First, several definitions will be elaborated that are important for understanding the research question. In addition, theories and models are examined based on existing literature that also contribute to answering the research question. This chapter consist primarily of the theory of walkability and the extent to which socio-demographic, such as socio-economic status (SES), and geographical determinants influence the extent to which people walk. Theories from the literature and policy data from the Dutch government have been combined to provide a clear overview of the current walking culture in the Netherlands. After explaining the term walkability in the Dutch context, the walkability index, on which this research builds, will be discussed. Including the way in which the current index is used, how it was developed and what the limits of this index are. Finally, a conceptual framework is illustrated to give a visual presentation of the research.

2.1 Literature review

Many recent as well as older studies have focused on the aspect of health in the city. They mainly focus on how citizens can be encouraged to adopt an active lifestyle. In many studies, it appears that the built environment and policy makers play a major role in this regard. For this research, relevant studies are examined that investigate how an active lifestyle is explained, why this is important, how the built environment plays a role, how citizens can be encouraged to adopt an active lifestyle and how this relates to practice. This chapter will address these themes and examine the findings of previous studies to understand the similarities and differences.

Active living and the built environment

In addition to individual choices and characteristics, the extent of the built environment has a significant influence on how far citizens can go in adopting an active lifestyle (Hansen et al., 2015, Haybatollahi et al.)

In recent years, much research has been conducted among others by Engelen et al. (2016), Hansen et al. (2015) and Sugiyama et al. (2008) on how the built environment influences an active lifestyle. The built environment determines to a large extent the travel options for citizens. This can be done, for example, by means of the presence or absence of public transport, the degree of greenery and the distance between the start and end of the journey (Haybatollahi et al., 2015). Most studies speak to a greater or lesser extent of the five Ds; "density", "diversity", "design", "destination accessibility" and "distance to transit" (Cervero & Kockelman, 1997, Haybatollahi et al., 2015, Vale & Pereira, 2016). These five indicators determine the extent to which citizens are encouraged to actively move through the built environment.

Besides these five indicators, there are several factors that influence an active lifestyle. For example, there is a large difference between rural and urban areas (Hansen et al., 2015, Cristina, Bouldin, & Battista, 2021) and even between different neighbourhoods within one city (Haybatollahi et al. 2015, Lau & Tan, 2018). This is partly due to the absence of a wide range of transport options and a long distance between living and working. As a result, people are not invited to use an active transport option, however, a difference can be seen between free time movements and compulsory movements (Song, Preston, & Brand, 2013). Song et al. (2013) have shown that an individual chooses to travel differently depending on the specific purpose, for example commuting or travel for leisure. People who travel for leisure more often tend to cycle or walk; people who commute more often tend to use 'faster' transport such as motorised transport. The type of target group therefore has a major influence on how active life is experienced (Sugiyama et al., 2008, Liao et al., 2020).

Walkability and the built environment
Studies from Song and Preston (2013), Faskunger (2013), Jun & Hur (2015), Singh (2016), Zuniga-Teran et al. (2017), Liao et al. (2020) and Wang and Yuqi (2019) acknowldge that the degree of active living is linked to the degree of walkability and cyclability. A place is walkable when all types of residents find it attractive, comfortable and safe to walk, whether for leisure or during their daily commute (Chapman, 1996). Therefore, walkable communities connect people to nature and destinations and honours social and environmental diversity. This leads to an improvement in safety, resource responsibility and social interaction and makes communities liveable and improves the quality of life (Cubukcu, 2013). In addition, the above studies show that a high degree of walkability and cyclability contributes to an active lifestyle. This degree is largely determined by the built environment and varies quite a bit between countries (Jun and Hur, 2015, Singh, 2016, Zuniga-Teran, et al., 2017, Liao et al., 2020, Wang and Yuqi, 2019).

All these studies are still quite recent and it appears that only since the 2000s, more attention has been given to research on walkability and cyclabilty than to research on, for example, cars and transport (Wu, Wang, Wang, Ta, & Chai, 2021). Many studies focus on the relationship between walkability and the built environment and how this affects the health of citizens, for example by looking at obesity. Hansen (2015) and Gell et al. (2015) state that an attractive neighbourhood with sidewalk cafes, density of landmark buildings and density of street trees leads to a lower BMI than places with none or less of these components. These places are less inviting to walk and therefore score worse in terms of walkability. As more studies have been conducted on walkability, more variables have emerged from the built environment that influence walkability. For example, most studies also mention separating roads, attractive greenery, addressing the human scale and orienting buildings to streets and open spaces (Grant, 2013, Lau & Tan, 2018).

Existing studies on walkability can be roughly divided into two categories. Studies that focus on health and studies that focus on the level of transport movements (planning). Both use the built environment as an indicator but from a different perspective. For example, health studies from Hansen et al. (2015) and Engelen et al. (2016) mainly focus on the negative effects of the existing built environment on health. Planning studies mainly focus on how to encourage people to use other transport options than using the car (Faskunger, 2013, Bajracharya, Too, & Khanjanasthiti, 2014, Sugiyama, 2008, Singh, 2016). It is sometimes difficult to compare the results from these studies because they are written from different starting points.

As several studies on walkability have been compiled and most of the variables have been uniformly assumed, several researchers have attempted to compile a walkability index. Leslie, et al. (2007) and Owen, et al. (2007) were among the first to develop a walkability index, which have been used in many subsequent studies (Wang and Yang, 2021). Liao et al. (2020) developed a walkability index for the Netherlands. However, this study has a number of limitations such as using or missing certain variables and no distinction was made in this study between different types of walks. These are aspects that will be included in this study. However, the study by Liao et al. (2020) will be used as a basis for this research.

2.2 Walkability

A place is walkable when all types of residents find it attractive, comfortable and safe to walk, whether for leisure or during their daily commute (Chapman, 1996). Nowadays that term has not really changed. Wang and Yang (2020) describe walkability as the degree to which the built environment invites people to walk, how it contributes to a healthy lifestyle and how it increases the quality of life for citizens. A place is therefore walkable when it invites people to walk instead of using motorized transport regardless if people are commuting or for leisure. The built environment is of great importance in the choice of transport of residents. In this research the concepts of GECCO will be used to determine if a place is walkable. These concepts are; Population density, Density of retail and service destinations (retail environment), Land use mix, Street connectivity (intersection density), Green space and Sidewalk presence (Wagtendonk & Lakerveld, 2019).

Jeff Speck, a renowned city planner and urban designer who wrote the book "Walkable City", describes 'A general theory of walkability' based on four elements that make an area more attractive for walking. A walk must fulfil four conditions; Firstly, it must be useful, that is, the aspects needed in daily life must be organised in such a way that they are within walking distance. Secondly, it must be safe; walkers must feel safe both socially and physically. Thirdly, it should be comfortable, people should be invited to go outside with a well decorated environment instead of large open spaces. Finally, it should be interesting to go outside, the built environment should radiate friendliness and humanity (Speck, 2012). Therefore, the built environment is one of the most important factors in achieving a correct walkability score. Speck mentions this as one of the four preconditions in his book.

However, many researchers have preceded him establishing the relationship between walkability and the built environment. To ensure a correct view of general and influential factors, the assessment of these elements requires a multidisciplinary study. Fitzsimons (2013) conducted one of the most general studies “A multidisciplinary study of walkability: its concept, assessment and applicability” on walkability and the built environment by reviewing about 27 books and scientific papers. He eventually classified walkability into two main groups: Functional environment and street scape. Functional environment consists of density, connectivity & permeability and land use and refers to “the structural, constructed environment that forms a city or town and its streetscapes”. Street scape refers to the visual elements of the like buildings, roads, sidewalks, green space and all other elements that influence the characteristics of the street.

All studies on walkability present roughly the same results, the difference being that different researchers use different factors or elements to determine the degree of walkability. Figure 2.1 shows the elements of walkability by Fitzsimons which is a general overview of many different studies on walkability combined into one scheme. This research will address walkability as a combination of Jeff Speck his view and the multidisciplinary research of Fitzsimons.

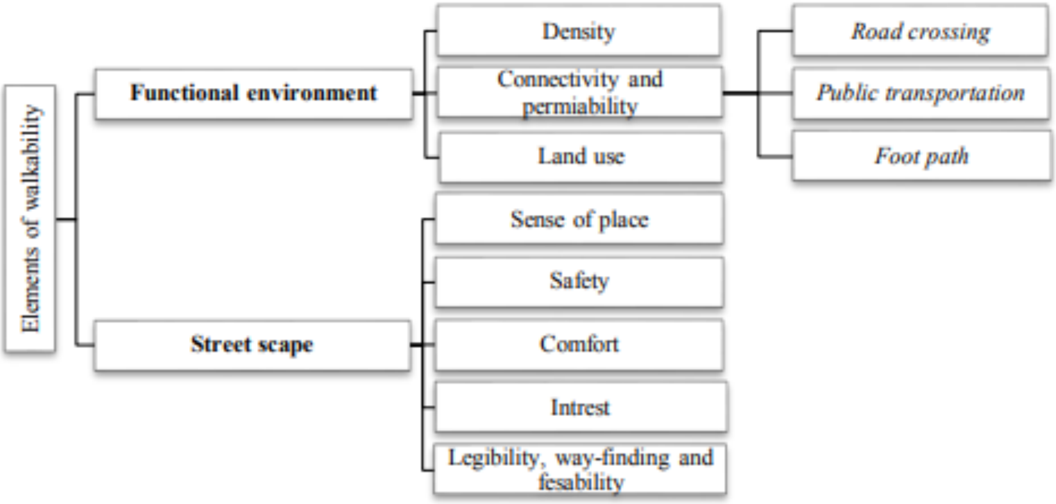


Figure 2.1. Correlated elements of walkability and the built environment (Fitzsimons, 2013)

2.3 Determinants of walkability in the Netherlands

The degree of walkability can be interpreted and determined in different ways. A distinction can be made between socio-demographic and geographical determinants that influence the degree of walkability in the Netherlands. Socio-demographic data can help to explain the use of different transport modes by individuals, these are mainly factors that are partly determined by the geographical determinants of an area. These geographical determinants differ per place resulting in different outcomes. For example, the built environment plays a major role in the degree of walkability and how individuals experience walkability in their neighbourhood (Rafiemanzelat, Emadi, & Kamali, 2017). In the result section of this research a link will be made between the results obtained from this research and the existing theory on walkability.

Internationally, the Netherlands is second only to Switzerland in the use of 'active modes', i.e. travel by bicycle or on foot. Half of all journeys, one tenth of all kilometres travelled (especially for short trips) and one third of the time spent on mobility are made by bicycle or on foot. The high international ranking of the Netherlands is mainly due to its high bicycle use; compared to other countries, the Dutch walk less than their fellow citizens. (Schaap, Harms, Kansen, & Wüst, 2015).

The motives for the Dutch walking policy are mainly recreational. In fact, one third of all pedestrian trips and one fifth of all bicycle trips have a recreational purpose. There is a clear difference between recreational and utilitarian purposes. Across the Netherlands, about 20% of the utilitarian trips are made on foot, in contrast to recreational trips where this share is almost 35% (CBS, 2010-2014). A later study from 2019 (de Haas & Hamersma, 2019) states that almost half of all walking trips in the Netherlands are derived from leisure. A quarter of the walking trips involve shopping, so walking for work is the least common. By investigating the distance covered on foot, it appears that almost two-third of these movements can be attributed to leisure time, so much less distance is covered for other walking motives (de Haas & Hamersma, 2019).

2.4 Built environment

Roof & Oleru (2008) describe the built environment as “the human-made space in which people live, work, recreate on a day-to-day basis including the buildings and spaces we create or modify”. More general, the built environment refers to manmade surroundings that provided by humans for human’s activity which include interdisciplinary elements (Rafiemanzelat, Emadi, & Kamali, 2017). For this research it is interesting to understand the built environment in relation to travel behaviour and the degree of walkability. The built environment in combination with travel behaviour can be described in five dimensions; Density, Diversity, Design, Destination accessibility and Distance to transit (Haybatollahi et al., 2015). These are the most spatial terms discribing the built envirnoment, each term having wider definitions. Think for example of the presence of reacreational and green spaces, safety, mixed land-use, residential and commercial density or well connected streets.

A high degree of walkability constitutes active living, thus active living is one of the core concepts for a healthy lifestyle and is strongly influenced by the built environment. The built environment in combination with fewer physical demands in working life and technological innovations have led to a more sedentary lifestyle, resulting in a treat to public health (Faskunger, 2013). An active lifestyle can be distinguished by two types of behaviour: physical activity and sedentary behaviour (Sugiyama, 2008). However, the one does not exclude the other; people who lead active lifestyles may still be forced to sit a lot in their daily activities, for example at work or due to the mode of transport. The built environment therefore still plays a major role and designing the built environment in the right way can help to reduce sedentary lifestyles and encourage active lifestyles. For example, by increasing the degree of walkability. An active lifestyle here refers mainly to how people are invited by the built environment to walk, cycle or take public transport in everyday life, both when commuting and for leisure.

In addition to the above-mentioned characteristics, there are other factors that contribute to the differences between geographical areas. For example, spatial planning plays a major role in the proportion of walking or cycling. These spatial characteristics are strongly linked to population characteristics. Bohte (2010) argues that the effect of self-selection plays an important role in this connection. People are more likely to choose where to live or not to live based on spatial characteristics, which explains why the population composition is partly influenced by the spatial characteristics of an area. These spatial characteristics, such as distance to facilities or the built environment, have a strong influence on the manner of moving. People who choose to live in a city centre live closer to amenities and have a greater opportunity to use 'active modes'. This is in contrast to, for example, rural areas with fewer amenities and a less designed infrastructure to encourage the use of active modes.

In a study of walkability, one of the most important factors is the user of the built environment. This involves identifying who uses the built environment, why they use the built environment and how this use is experienced by different individuals. Hiller & Leaman (1973) and Lang (1987) state that user centred theories are situated on a continuum that varies between, on the one hand, a deterministic definition of the relationship between the environment and behaviour and, on the other hand, the influence of the built environment on users (Vischer, 2008).

Figure 2.2 shows, on the one hand, the cause-and-effect perspective that assumes that the environment in which people find themselves, and how it is designed, determines how people behave in that area. Identifying the behaviour of people that is influenced by the built environment is called 'environmental determinism'. At the other end of the continuum is social constructivism. This way of thinking assumes that reality and human experience are entirely socially constructed and thus not influenced by the physical environment in which an individual finds himself (Mead, 1962).

Figure 2.2. The continuum of user centred theories (Mead, 1962).

Although these extremes are not realistic in today's context, every user of the built environment will find himself somewhere between the extremes on the continuum. Behaviour is indeed influenced by the built environment but, conversely, not all behaviour is caused by the built environment. The social context, feelings, intentions and attitudes are at least as decisive for the behaviour shown by individuals in the built environment (Vischer, 2008). This research will therefore consider the different positions of individuals on this continuum. Using the LifeLines data, which collects life aspects in the broadest possible way, it can be determined to what extent the built environment influences behaviour and how the built environment is experienced by individuals.

The built environment also includes the degree and the design of the infrastructure which also plays a role within walkable cities. The degree of mobility policy, parking tariffs and the design of the infrastructure all play a role in the extent to which inhabitants are invited to use 'active modes'. Think of the quantity and quality of the walking and cycling infrastructure, but also the extent to which car use is discouraged and the mobility policy in general. Policy therefore determines, as do demographic and geographical characteristics, the extent to which inhabitants are encouraged to use 'active modes' which determines their travel behaviour.

Bhat and Guo (2005) state that the relationship between the built environment and travel behaviour is very complex. This can be shown by three elements that reflect this complex relationship. First, the multidimensional nature: both the built environment and travel behaviour are multidimensional in nature. This means that for both concepts, many different dimensions exist, and it is not always clear which ones relate to each other. As a result, they are different in each context, making it difficult to make a homogeneous comparison. Secondly, the moderating influence of decision-maker characteristics. These characteristics can be, for example, socio-demographic, travel-related or environmental and can directly or indirectly influence the relationship. The perception that is chosen then already strongly determines the outcome of a study. Third, the spatial scale of analysis to measure interventions in the built environment. Most studies use predefined spatial data such as postal codes, neighbourhoods or municipalities. This is because these data are easy to compare with, for example, transportation data. However, in reality, citizens may experience the space and scale of a neighbourhood differently and make choices based on which spatial characteristics they choose. Besides travel behaviour, there is also the issue of self-selection; Do active environments produce active inhabitants, or do active inhabitants seek active environments? Not everyone is capable of living where they would like to live due to their socio-demographic or economic circumstances.

2.5 Walkability index

The walkability index used in this study was prepared by the Geoscience and health cohort consortium (GECCO). GECCO is a Dutch infrastructure to support research to study the relation between environmental characteristics and health (GECCO, 2022). Their index is based on the following components: Population density, Density of retail and service destinations (retail environment), Land use mix, Street connectivity (intersection density), Green space and Sidewalk presence. GECCO itself devised these factors based on scientific literature. In addition, they found other factors such as safety from crime, traffic safety, traffic volume and speed, pedestrian crossing availability, aesthetics, air quality, shade or sun, street furniture, wind conditions, job density and specific walking destinations such as bus stops (Wagtendonk & Lakerveld, 2019). In order not to complicate the analysis more, these factors have been omitted or combined into the six above mentioned factors, that according to GECCO, can produce a valid index.

Like this research, the main objective of GECCO's research was, besides developing a walkability index, to relate the walkability scores to health behaviours and health outcomes in a potentially wide range of epidemiological research (Wagtendonk & Lakerveld, 2019). Important to understand in GECCO's study is that walkability scores derive from recorded physical activities by respondents. From these respondents only the residential postal codes or address coordinates are known and thus available for analysis. This ensures that the index can only provide information about trips from the respondent's home to a certain location. A walk from their work to the supermarket or a walking trip for leisure further away cannot be measured in this index as the respondent's address is the only known geographical unit. Because such a walkability cannot address multiple epidemiological research goals and pedestrian types at the same time, the developed index is based on a set of components or factors for different walking ranges. Combined, these factors can produce a score between 0-100 (low to high walkability) to develop an index that meets a specific objective (Wagtendonk & Lakerveld, 2019). To understand how the walkability index of GECCO is set up, the corresponding technical document is attached to the final submission.

2.6 Limitations to the walkability index

The degree to which an area receives a high or low walkability score can give a distorted impression depending on the location. Current walkability indexes stick to geographically delimited areas. This is because it is easier to extract and compare data based on geographical determined areas such as postal code areas (PC6) than on, for example, single dwellings. Furthermore, when conducting epidemiological research, privacy is highly regarded which means that response on questionnaires cannot be traced back to the respondents. This means that the lowest scale possible are geographical units such as postal codes. The problem this creates is that there is a big difference between postal code areas in the city (urban areas) and in villages or the countryside (rural areas). A postal code area in the city of Groningen, for example, is quite detailed and sometimes only consists of one or a partial street. In rural areas outside the city, on the other hand, an area hundred times the size may contain only two or three farms. GECCO uses the centre of a postal code area (centroid) to determine the WI-score for that PC6 area. However, the centre of a large PC6 area is not representative for that whole area and thus the walkability score then usually gives a misleading result.

A good example can be seen in figure 2.3 and figure 2.4. These images show the grid of green space, one of the components used to create the walkability index. GECCO bases the WI-score for a PC6 area on a centroid, which means that the grid value below the centroid determines the value for the whole PC6 area. In the example shown in figure 2.2, the value of the blue PC6 area for green space would be about 90. The higher the value of green space, the higher the score in the WI will be. The problem that arises from this approach is that (especially with large PC6 areas) a distorted result is created. In this PC6 area, although a large part of the area contains a green environment, the western part of this PC6 area has a lower score, between 0 and 90. A score of 90 based on green space would not be accurate for this area since about 30% of the total area has a green space value of less than 90.

This problem occurs to a lesser extent in urban PC6 areas. These areas are often considerably smaller and this entails that the value of the centroid more often matches the values around the centroid. In addition, a buffer of 500 metres is calculated from each location to determine its values. In an area that has a smaller surface area than a 500-metre buffer, it therefore makes little difference which value is used since all the values are approximately the same. Figure 2.3 shows a PC6 area in an urban area, here the centroid approach is less inaccurate.

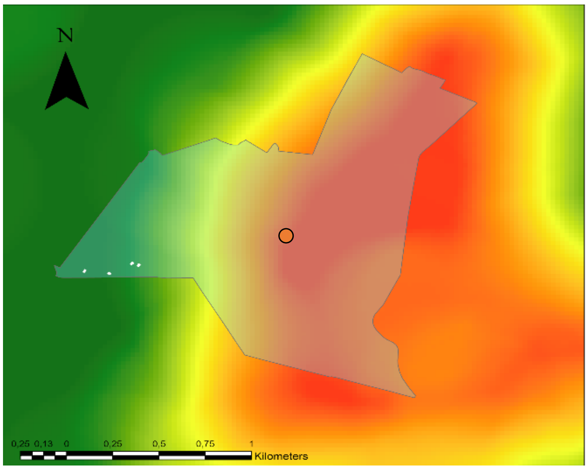


Figure 2.2. Centroid approach urban PC6 area

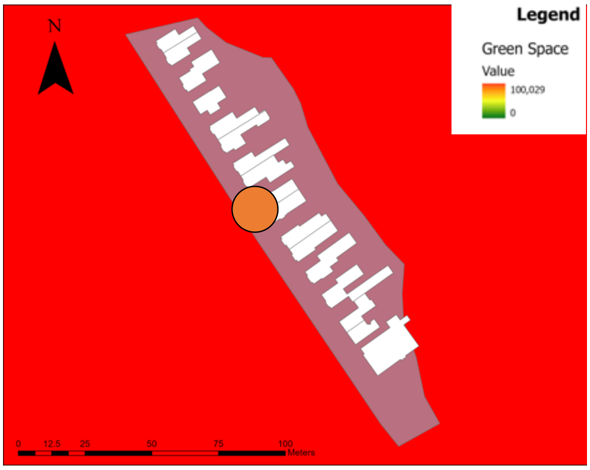


Figure 2.3. Centroid approach rural PC6 area

To overcome this problem, this study uses different ways to minimise this margin of inaccuracy. A first approach that is used is that of taking the average grid value of the whole PC6 area, this is also referred to as the *PC6-approach* in this research. This calculation includes all values in the entire PC6 area rather than just the value under the centroid. If this method were applied to the PC6 area in Figure 2.3, both high and low values for green space would be included in the calculation. The average of these values is then used to determine the value for the whole PC6 area. The PC6-approach will therefore be even more accurate for urban areas (figure 2.4) because now the values in the entire area will be included. However, this method is probably still inaccurate for large PC6 areas where relatively few people live

However, there is a chance that also this approach is not accurate enough in rural areas when the index would be used for health research. Figure 2.3 shows that a total of four houses lie in the west of the PC6 area, which are the only houses in this PC6 area. If someone in this area would like to know the WI-score for his property, it is given based on the value of the entire PC6 area. Regardless of someone lives in the west or in the east of that area, all areas in that PC6 area will receive the same value. To minimise this problem, a second approach will be used. In this approach, average values are assumed, but only for those grid values that are below a residential property. This method is called the *Residential Property approach* (RP) in this research.

Both methods are explained in more detail in the method section, with an explanation of how they were created and to what extent they are easy to implement and feasible for follow-up studies. Especially the last method will be time-consuming and use a lot of computing power. The results of both methods are further explained in the results section and compared to the Centroid-approach. This study tries to filter out the margins of error and to create a better way of developing a walkability index that is representative for both rural and urban areas. To distinguish between urban and rural, and especially between different rural areas, a new classification will be set up with better distinction between these areas.

2.8 Conceptual Framework

The conceptual framework illustrates how the different variables relate to each other. In this study, the built environment and people's behaviour determine an active lifestyle. Firstly, the built environment, the theories of Speck and Bhat and Guo show that behaviour is driven by the built environment in which an individual finds himself. On the contrary, Hiller & Leaman (1973) and Lang (1987) state that human experience is entirely socially constructed and rely on rational actions without being influenced by the built environment. Active living is thus determined by a relationship between behaviour and the built environment.

The extent to which the built environment does or does not contribute to an active lifestyle can be determined by examining the walkability score of that area. In this research the walkability index from GECCO will used. GECCO based this index on six components: population density, density of retail and service destinations (retail environment), land use mix, street connectivity (intersection density), green space and sidewalk presence.

Because behaviour and the built environment are not the same everywhere and depend on where an individual lives, a mediating variable is added, namely whether the individual lives in an urban or rural society. In urban, metropolitan areas, one may be more likely to be encouraged to live actively because buildings are closer together and necessities of life are more concentrated than in rural areas.

Finally, the LifeLines data is added to show individuals active travel behaviour, where these people live and how they experience the built environment.

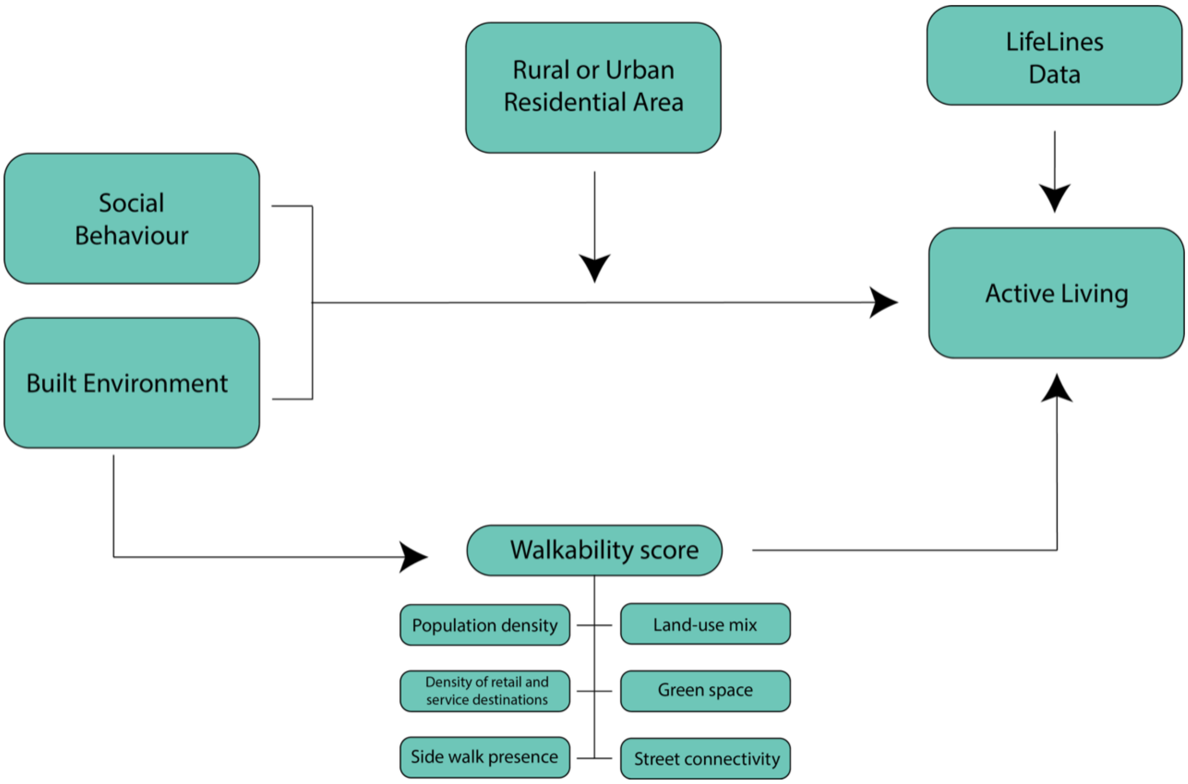


Figure 2.4. Conceptual framework

3. Methodology

In order to perform valid research, a well-designed methodology section is necessary in which the design of the research will be explained. In this chapter the research methods and the research strategy are being discussed to help understanding the main research question. First, the research philosophy, or research paradigm, that determines this research will be elaborated. This will be followed by the research methods and the optimization of these methods by explaining the applied qualitative and quantitative methods. Also, the collection of data and the principles of reliability and validity will be discussed in this chapter.

3.1 Research philosophy

The methods and approach are mainly determined by the research philosophy, or research paradigms. A paradigm can be described as a world view, or as Guba and Lincoln (1990) describe it; "A basic set of beliefs that guide action". The chosen way of thinking partly determines the design of the research.

The main part of this research focuses on improving the walkability index fur rural areas and to what extent this contributes to a better analysis and implementation of the LifeLines dataset. In this part of the research, Geographical Information Systems (GIS) are mainly used. GIS is a useful tool for spatial questions where large amounts of data can be merged, analysed and calculated with different data sources. The GIS research complements the health research carried out at Radboud UMC, where new insights can be gained in combination with statistical programs. The use of GIS and health studies by a researcher is post-positivist in nature. Post-positivists acknowledge that theories, background knowledge and values of the researcher may influence what is observed (Robson, 2015). Guba and Linclon (1994) state that "Reality is assumed to exist but to be only imperfectly apprehendable because of basically flawed human intellectual mechanisms and the fundamentally intractable nature of human phenomena". That is because in GIS-studies, quantitative data is analysed in which the results are always approached objectively by the researcher. No geographical model predicts human behaviour in the way that it does in reality. The challenge is therefore to be able to interpret and process spatial data in a GIS model as reliably as possible, although this will inevitably lead to some form of inaccuracy.

The last part of this research is combining the adjusted walkability index with the LifeLines data obtained from questionnaires over 30 years. Working with surveys about healthcare and other social issues means that a subjective approach to social pursuits must be considered since it is focused on an individual's perception of a phenomenon. This part of the research is of social constructivist nature, social constructivism assumes that experiences of individuals are partly constructed by social processes depending on the society an individual finds himself (Kukla, 2000).

3.2 Research strategy and methods

The research strategy is the overall design of this study which includes different methods of data collection and analysis. This research consists of qualitative and quantitative methods to gather data which means there isn't just one research strategy. Van Thiel (2007) discussed four main research strategies. Namely, the case study, the experiment, the desk research and the survey, all four strategies consist of different methods and techniques. In her book, van Thiel discusses the following methods; observation, questionnaires, interviews, content analysis, meta-analysis and secondary analysis.

The strategy that fits best to this research is desk research. Most of the data being used has been collected or produced by someone else. In this research, the policy documents are written by government agencies who provide Dutch walking data. The walkability index is created by GECCO and the LifeLines surveys are held by LifeLines. This means that almost all used data sources comes from secondary data.

A problem that could arise is that these data sources where originally produced for another purpose which means that the data will not always match with the research question. It is therefore necessary that in this research information must be found that matches the main question as adequately as possible (Van Thiel, 2007).

Van Thiel describes three main methods for analysing existing data. Namely, content analysis, meta-analysis and secondary analysis. These ways of analysis consist of different ways of collecting and processing data.

The first sub-question; "What constitutes walkability in the Dutch context?" will be answered by means of a literature review or in other words; content analysis. This literature study investigates the Dutch walking culture. The literature review on walking in the Netherlands is conducted to support the subsequent geographical analyses, and the literature also supports the interpretation of the results that emerge from these analyses. This literature study uses qualitative data; among other things, mobility documents from the Dutch government and Dutch studies into walkability are consulted in order to provide a description of the Dutch walking culture. Furthermore, a statement can be made about the current level of walkability in the Netherlands and the Dutch walking culture. With the help of the literature study, the subject will be delineated, and a clear definition of the term walkability in the Dutch context can be given.

To subsequently work with these findings and to answer the main question and sub question two and three, quantitative data will be used in cross-sectional research. The data is examined at a point in time, this allows a statement to be made without intervening or randomizing (Stilma & Rijkenberg, 2021). By using a GIS, the current index will be evaluated and remodelled to a better geographical unit for rural areas. Furthermore, statistical programme's such as R-Studio and SPSS in combination with GIS will help to analyse relations between the walkability scores and individual's active mobility behaviour. This resembles secondary analysis; existing numerical data is being analysed. Existing data sources are being combined after which a new dataset is the result. Secondary analysis involves statistical data and is mostly useful for deductive or hypothesis-testing forms of research, in this case combining and analysing statistical with GIS data. The problem that arises with this technique is optimization of the data. Using different types of data sources means different outputs. To combine and analyse both datasets, the data has to be optimized or recalculated which could lead to data losses. To overcome this problem all adjustments have to be documented and attention must be given to the reliability and validity.

This research thus combines different strategies and methods to answer the main question. John Creswell states that "Mixed methods research is a research design (or methodology) in which the researcher collects, analyses, and mixes (integrates or connects) both quantitative and qualitative data in a single study or a multiphase program of inquiry" (Johnson, Onwuegbuzie, & Turner, 2007) This research therefore uses the three main research paradigms and different methods of desk research. Mixed methods is the most useful method to work on this research because it allows combining qualitative and quantitative data which is ultimately the best approach to answer the main question as the quantitative results can be supported by qualitative data.

3.3 GIS Analysis

The main part of this research consists of geographical analyses. Sub questions two and three are answered by means of a Geographic Information System, or GIS. In a GIS, spatial datasets can be loaded, visualised, and analysed with and between different datasets. This way, large spatial issues can easily be analysed. In this research, the software program ArcGIS is used. The choice for this software is based on the extensive functionalities and simple integration with statistical programs that ArcGIS offers, compared to other (free) software programs.
As mentioned in the theoretical framework, a walkability index has been drawn up by GECCO. Afterwards, it appeared that this index is mainly suitable for urban areas because in these areas the PC6 areas are smaller than in rural areas. A large PC6 area with relatively few inhabitants quickly gives a distorted representation of the actual situation. To solve this problem, several options have been devised to enable a more accurate calculation to be made in rural areas, which are explained in this section.

GECCO calculated the walkability index using seven components in GIS. These different components are computed as grid values with grid cells of 25x25 metre covering the Netherlands in a regular spaced grid. The individual grid cell values for each component are computed using focal statistics (spatial statistics / summaries) for different years and nine different sized buffers around each grid cell (Wagtendonk & Lakerveld, 2019).

"*The result of the GIS operations is a set of tables in which for each residential address location in the Netherlands a raster value is available for each of the seven walkability components. All component values are converted in z-scores (representing the number of standard deviations from the population mean) before summing them. The combination and computation into a total walkability score for each selected year, each location and buffer size are carried out within SPSS using the following formula: Walkability = (z-population density) + (z-street connectivity) + (z-land use mix) + (z-retail environment) + (z-green space) + (z-pedestrian area density) Finally, the score is normalized to derive values for the walkability index between 0 and 100. The higher the score, the greater the ability to walk in that area."* (Wagtendonk & Lakerveld, 2019)

To determine the value of a component in one PC6 area, GECCO calculates with so-called centroids. A centroid is the centre of a PC6 area, and this centre point (calculated in GIS) determines the WI-value of that entire PC6 area. In small PC6 areas this gives a reasonably accurate input for the value determination. However, in large PC6 areas the centre of such an area can have a totally different value than the outskirts of this area. GECCO accepted this margin of inaccuracy as other methods would be time-consuming and, according to them, this did not have a major impact on the overall walkability index. However, for LifeLines this is a problem since their respondents are located in the provinces Friesland, Groningen and Drenthe. These are provinces with relatively many rural areas and therefore also large PC6 areas. To solve this level of inaccuracy, two techniques are used in this research that contribute to a better approximation of the walkability scores in rural PC6 areas. These techniques help to answer sub question two: *What geographical unit is most suitable for conducting walkability studies to validate the index for all geographical areas in the Netherlands?*

Average of PC6 area (PC6-approach)

The first approach used is to calculate the average grid values in the whole PC6 area. In other words, instead of only using the value of the centre of the PC6 area (centroid), in this approach the average is taken of all the grid values in the relevant PC6 area, thus reducing the margin of inaccuracy. For example, in a large PC6 area, the values in the east of the PC6 area may be much higher than the values in the west but this approach ensures that the values of the PC6 area are determined by taking the average of all the grid values in that PC6 area. To perform this calculation, it is first necessary to plot the PC6 areas of LifeLines on the map and import the raw grid files of the seven components. Then, using the analysis function "zonal statistics to table", the average value of a grid file under a given PC6 area is calculated and displayed in a table. This table is the output of this analysis and contains the PC6 areas and the average grid value belonging to that PC6 area. This table can then be "joined" to the original PC6 file. The join function checks which PC6 area from the output table matches with a PC6 area from the original PC6 file. This is necessary as a table cannot be geo-referenced, so the table must be linked to a spatial dataset, in this case the original PC6 file. After the join is executed, the whole file is exported as an Excel table. These steps are carried out for all seven components, which results in one comprehensible table containing the average grid value for each PC6 area. Figure 3.1 shows a PC6 area and the grid of the green pace component. To calculate the value of the whole PC6 area, all grid values in a PC6 area will be summed up and then divided by the number of grid cells to determine the average score of green space for this PC6 area.

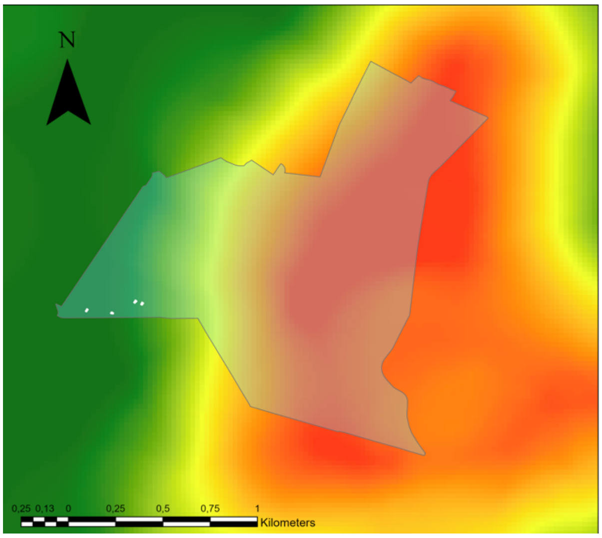


Figure 3.1. PC6 approach rural PC6 area

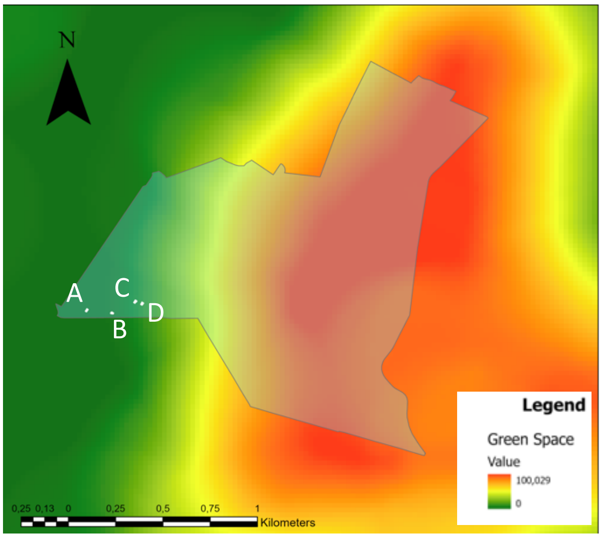


Figure 3.2. RP approach rural PC6 area

Average of properties (RP-approach)

The approach of calculating a value using the average grid values under the PC6 area seems to be a better approach than using a centroid. However, there is still a margin of inaccuracy. This is because large PC6 areas often contain only a few dwellings, whereas a walkability index assumes places where people live. Suppose a PC6 area with a surface area of one hundred football fields contains a total of four farms spread across the whole area. Although the previous approach took the average of the whole PC6 area, there can still be large differences between the different farms. Figure 3.2 shows the four residential properties in this area in the west of the PC6 area. In this example, farm A has a green space value of 0.5, farm B a value of 3, farm C a value of 8 and farm D a value of 11. If somewhere else in the PC6 area extremely high or low values occur, the average of this area will not match the average values of the farms. To minimise this margin of error, this approach assumes the average grid values under buildings where people live, in other words, residential properties. By not assuming the average grid value of the entire PC6 area but only the average grid value among residential properties, a more realistic outcome is outlined with respect to walkability.

In the example of the four farms, the value of the PC6 area will then be (0.5 + 3 + 8 + 11) / 4 = 5.6 instead of an average value from all the grid values of the whole area. This approach seems to be the most accurate of the three because it assumes places where people live. A walkability index is based on the behaviour of people and areas where no people live are therefore less relevant for such an index, especially when the index is built up from components such as population density, sidewalk presence or land use mix.

The calculation for this approach is similar to the previous calculation. For this analysis, it is also necessary to import the PC6 areas and the raw grid values. However, in order to calculate the values for residential properties, a dataset of all buildings in the Netherlands is also required. This dataset can be retrieved from the 'Basisregistratie Adressen en Gebouwen' (BAG). This data set is made public by the Dutch government and is therefore accessible to everyone. The BAG dataset contains all buildings in the Netherlands with the corresponding location and function. For this calculation, however, only the residential properties are required, so these must be filtered out. Next, only the residential properties that are located in a PC6 area where respondents of LifeLines live are needed. A “clip" between the residential properties and the PC6 areas of LifeLines leaves only the residential properties necessary for the analysis. It is advisable to do this first as the BAG file is so large that calculating with this entire file is very time-consuming. Then, as in the previous approach, the "zonal statistics to table" function is used. This analysis then shows all the grid values under the selected residential objects. Note that the cell value is reduced, as the programme initially calculates with the original cell value of 25 by 25 metres. However, because most residential properties are much smaller, not all properties are included in the calculation. In this analysis, therefore, a cell value of 0.5 metres by 0.5 metres was chosen. This results in analyses that take a lot of time but ensures that virtually all properties are included in the analysis. Next, the output table has to be joined again, this time with the residential objects file because a table cannot be geo-referenced. After the join, each residential object in the BAG file has the corresponding underlying grid value. As explained in the theoretical framework, only PC6 areas can be used, which means that the values of the residential objects must be converted to PC6 areas. This is done using the "Spatial Join" function whereby attributes from one feature are joined to another feature based on the spatial relationship. The target feature (PC6 area) and the attributes from the join feature (residential objects) are written to the output feature class. The newly created feature class consists of PC6 areas and the average values among the residential objects in that particular PC6 area. This analysis is carried out for all seven components and then exported to Excel where it can be used to create a conveniently arranged table containing the average grid value per component and per PC6 area.

After calculating the new WI of both methods, it is interesting to find out which approach explains the WI better. In statistical programs, this can be compared using regression models but could also be investigated visually in ArcGIS Pro. The PC6 areas that have the greatest mutual difference in the WI between the two approaches are then compared. Using the raw grid files, it can be determined which approach is more accurate in computing a WI-score, especially for rural areas.

3.4 SPSS & R-Studio analysis

After both Excel tables (average of PC6 and average of residential properties) have been exported, they will be imported into SPSS. SPSS is a statistics program that can easily perform statistical calculations using the export tables from ArcGIS. In SPSS, the tables are compared to determine if significant differences or correlations between the two different approaches can be found. In this way, it can be validated whether the second approach is more or less accurate than the first one.

First, both walkability indexes are uploaded as a table. The table consists of 18 columns and 36.373 rows. Each row represents one PC6 area. The names of the columns are listed in table 3.1:

PC6	Opp	OAD	STED	GS6	GSP	LM6	LMP	PD6	PDP	PT6	PTP	SC6	SCP	SW6	SWP	RS6	RSP

Abbreviation	Description
PC6	Postcode six area
Opp	Area of PC6 area
OAD	“Adressenomgevingsdichtheid” (Surrounding address density) based on CBS data
STED	Degree of urbanisation based on CBS data
GS6	Green Space score of the PC6 approach
GSP	Green Space score from the Residential Properties approach
LM6	Land Use Mix score from the PC6 approach
LMP	Land Use Mix score from the Residential Properties approach
PD6	Public Density score of the PC6 approach
PDP	Public Density score of the Residential Properties approach
PT6	Public Transport Density score of the PC6 approach
PTP	Public Transport Density score of the Residential Properties approach
SC6	Street connectivity score of the PC6 approach
SCP	Street Connectivity score of the Residential Properties approach
SW6	Sidewalk Presence score of the PC6 approach
SWP	Sidewalk Presence score of the Residential Properties approach
RS6	Density of Retail and Service Destinations score of the PC6 approach
RSP	Density of Retail and Service score of the Residential Properties approach

Table 3.1. Column names of Excel table with WI-scores of PC6- and RP-approaches combined

Each component has a certain mean value depending on the corresponding PC6 area. Based on this table, two new Walkability indices will be created. This is realised by first standardising the component values to Z-values. Standardising statistical observations is an artefact to correct unequal distributions between groups according to certain characteristics using a weighting. (Pelfrene, 2010). SPSS calculates the Z-score by dividing the difference between the mean-value by the standard deviation (IBM, 2020), these Z-values are summed up and converted into a formula. In this formula, the lowest component value is subtracted from the chosen component value and divided by the highest component value minus the lowest component value. In a formula, it looks like this:

100 * ((value component - Min. value component) / (Max. value component - Min. value Component))

This formula is executed for both the PC6-approach and the RP-approach after which each unique PC6 area eventually obtains a certain WI-score. After performing these calculations, regression analyses can be used to calculate if a significant difference has occurred and which correlations can be found between both approaches.

Because this research mainly focused on rural areas, the emphasis will be on those areas with a low degree of urbanisation compared to areas with a very high degree of urbanisation.
The urbanity is measured by the "Omgevingsadressendichtheid" (OAD) or in English; Surrounding address density, which is compiled by the CBS. "The OAD is expressed in addresses per square kilometre and aims to reflect the degree of concentration of human activities (living, working, going to school, shopping, going out etc)" (CBS, 2022). These data are available at PC5 level and can therefore easily be converted to PC6 areas. Based on the OAD, the degree of urbanisation (DoU) for a particular area can be determined. CBS distinguishes between five classes:

1. Very Urban: average oad of 2500 or more addresses per km2;
2. Strongly urban: average oad of 1500 to 2500 addresses per km2;
3. Moderately urban: average oad of 1000 to 1500 addresses per km2;
4. Low urban: average oad of 500 to 1000 addresses per km2;
5. Non-urban: average oad of less than 500 addresses per km2.

This research mainly focuses on the degree of urbanisation of one and five. One is therefore highly urban (urban) and five non-urban (or rural). The analyses are done by subdividing the PC6 areas into one of the five classes, so that comparisons can be made more easily between urban and rural areas.

Using the degree of urbanisation, the Centroid-approach, the PC6-approach, and the RP-approach are compared. First, the Centroid-approach is compared with the PC6-approach in a scatterplot to illustrate the mutual correlation. The same is done for the Centroid- and the RP-approach. In doing so, it is possible to compare which approach correlates strongly with the Centroid-approach. Next, the PC6-approach and RP-approach are compared to see how much the two approaches differ.

To understand the proportion of rural areas in de Netherlands, figure 3.3 illustrates the DoU in the Netherlands. The map clearly shows that a large part of the Netherlands falls into the first category with a DoU of less than 500 inhabitants per km/2.

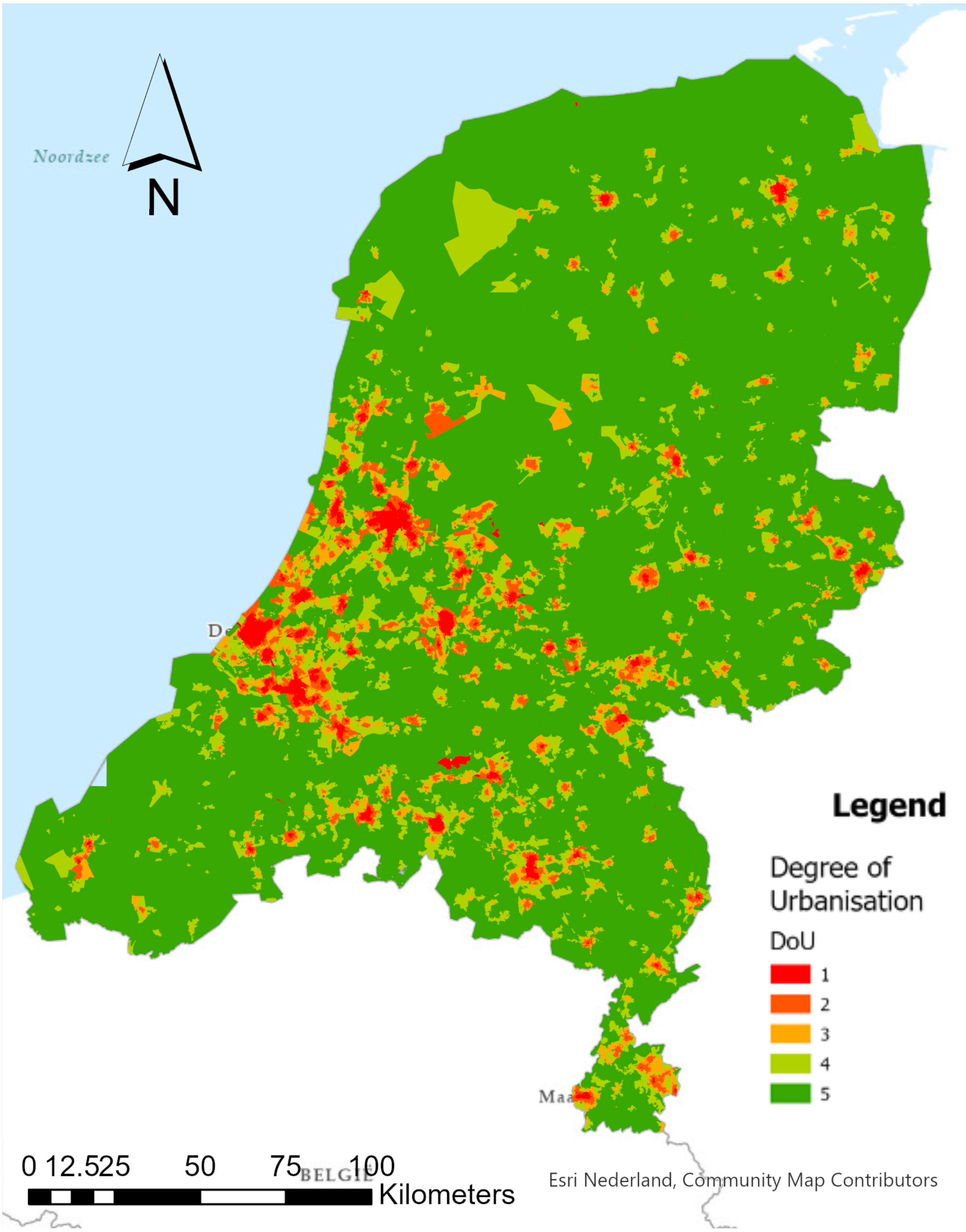


Figure 3.3. Degree of urbanisation in the Netherlands

3.5 LifeLines

In the final part of this study, the LifeLines data will be compared with the new WI to understand if the new approaches predict the WI for the rural PC6 areas better than the Centroid-approach. The first sub-question investigated, with the aid of literature research, what the Dutch walking culture looks like and which socio-demographic factors influence walking in the Netherlands. Comparable socio-demographic data is retrieved from the LifeLines database in order to make a comparison with the theory in the literature and the practical reality of the LifeLines participants. The aim of this comparison is to examine if the two new approaches result in a better WI-score than the Centroid-approach does. These results will be compared with the literature to determine if one or both approaches is better.

To perform these analysis, the LifeLines data must be requested in advance because their dataset is in a secure environment. The factors from the dataset that will be requested to make these comparisons are 'Walking time' (the walking time is compared with the seven components and the degree of urbanisation to show what a change of one of these components brings in the walking time to work or school). In addition, age, gender, educational level, employment situation and income are compared across all three WI. The aim of these comparisons is to show if a difference arises when a particular WI-approach is used compared to the other WI-approaches. In addition to socio-demographic factors, geographical factors are also examined for the same purpose. The aim of these comparisons is to answer sub question 3; *Which relations can be found between the walkability scores and individuals active mobility behaviour in different neighbourhoods and between different target groups?* Ultimately this, together with sub-question one and two, contributes to answering the main question; *To what extent can we evaluate and remodel the existing walkability index from GECCO to provide a more valid and reliable index?*

3.6 Data collection

In this study, various data are used to answer the main and sub-questions. In this paragraph, the realisation of the data sources used and how they are applied in this research are explained per sub-question.

To answer the first sub-question, what is the Dutch walking culture like, mainly literature will be consulted by means of desk research in which a distinction will be made between urban and rural areas. To understand the Dutch walking culture mainly Dutch policy documents on mobility will be examined.

To answer sub-question two, which geographical unit is most suitable for a walkability index, the current classification, namely PC6 areas, will be used first. These PC6 areas will be obtained from the CBS's "Kerncijfers per postcode". This is a dataset of all PC4, 5 and 6 areas in the Netherlands with additional detailed information. In this research the PC6 areas of 2018 will be used as the walkability index data from GECCO also originated from 2018. To calculate the new components scores, the raw grid values of all seven components of GECCO will be used, these are obtained from GECCO. For calculating the PC6-approach, only the raw grid values and the PC6 areas of 2018 are necessary. To calculate the RP-approach, all residential properties of the northern parts of the Netherlands are required. These can be obtained from the BAG, a large dataset with all buildings in the Netherlands including additional data, for example the function of the building. In this research, only those buildings with a residential function will be used.

To answer sub-question three, the newly created WI will be compared with the LifeLines dataset. The LifeLines data contains questionnaires and corresponding answers of all participants including the PC6 area they inhabit. The LifeLines data is secured and can only be retrieved from LifeLines, working with this data means that a request must be submitted to LifeLines for using their data. From their dataset mainly socio-demographic data will be used for comparing that data with both Walkablity indexes and the theory of sub-question one.

Name of dataset	Name in thesis	Source	Year
PC6zonalstat18_7components_500_1650_raw.csv	Raw grid values	GECCO	2018
CBS_PC6_2018-v3.zip§	PC6 areas	Cbs.nl	2018
BAG (EPSG:28992) Geopackage	BAG	Pdok.nl	2022
LifeLines dataset	LifeLines data	LifeLines	2019

Table 3.2. Used datasets

3.7 Validity

To determine internal validity, it is necessary to assess whether the instrument used in this study is fit for purpose. This research compares geographical data with socio-demographic data. For measuring and analysing geographical data, the use of GIS is the best method. In a GIS, large amounts of spatial data can be merged and compared with other spatial data. There are several GIS programmes, such as ArcGIS Pro and Q-GIS. In this research ArcGIS Pro will be used because Q-GIS has less functionalities than ArcGIS Pro. Comparing spatial data and socio-demographic data can be done both in GIS and in statistical programmes such as R-studio and SPSS. R-studio has a good integration with GIS system but SPSS is more user-friendly so SPSS will mainly be used. External validity is determined by determining if the research results are generalisable, i.e. whether the results can be applied to the entire population. This study uses data from 165,000 inhabitants of the Northern Netherlands in all age groups. Because this is a large varying group, the results can be generalised to the rest of the Netherlands. Although fewer people live in the Northern Netherlands than in other parts of the country, there still is a distinction between urban and rural areas comparable to the rest of the

Netherlands, except the Randstad. However, this research mainly tries to solve inaccuracy for rural areas many of which can be found in the Northern Netherlands.

A first point of attention is the way in which LifeLines participants gave their information. The problem of social-desired answers or 'response errors' (people misunderstand the question) should always be considered (Van Thiel, 2007). Furthermore, the original walkability index by GECCO was based on scientific literature, mainly by studies form the USA. In the USA, factors that determine walkability differ from European or Dutch areas, this means that maybe not al components apply well to the Dutch context.

3.8 Reliability

The reliability of a survey must be found in the degree to which the measuring instruments are reliable (Van Thiel, 2007). In other words, are the results accurate and reproducible. This research uses geographic and socio-demographic data that are fixed, these data cannot change. The analyses carried out are all statistical in nature and, assuming they are carried out properly, the same analyses with the same data will give the same results. Furthermore, because this research also focusses on health-related data, choices will be discussed with researchers form the Radboud UMC to enhance 'inter research reliability' (Van Thiel, 2007).

This research is more technical research due to the analysis in GIS and SPSS. Using GIS as main method ensures that all steps in the analysis can be reproduced. In appendix I, a technical document can be found which describes all analysis executed in this research, step-by-step. Doing so, someone with no experience in GIS could perform the same analysis with the same datasets, as all data is online available, except the LifeLines dataset. however, all WI-approaches can be reproduced with different datasets, it is therefore not necessary to use the LifeLines dataset for calculating the Centroid- PC6- or RP-approaches. Also, the analysis in SPSS can easily be reproduced as the input dataset derived from the output file from the analysis in GIS.

4. Research Results

In this chapter, the results generated by the GIS and SPSS analyses are presented, explained and analysed. This is done in the same order as in which the sub-questions were formulated. In the theoretical framework, walking figures from mobility documents of the Dutch government are discussed, interpreted and compared with data from the analyses conducted in this chapter. In addition, the PC6-approach analysis is compared to the RP-approach analysis to determine if any differences or similarities can be observed. Finally, the Walkability indexes are compared with the LifeLines data, in order to evaluate the results against the data from the literature. Ultimately, a conclusion can be drawn as to whether the new analysis approaches are useful for integrating geographical data with health data or if no significant difference can be found.

4.1 Dutch walking culture

The Dutch walk relatively much compared to other countries. The highest proportion of these trips consists of walking for leisure, followed by walking for relaxation such as shopping and subsequently walking for education. The Dutch do not walk often to work, this is mainly due to the larger distance between living and working in the Netherlands and the willingness to walk, this is measured by the acceptable walking distance. It appeared that different situations generate different acceptable walking distances. In this results section, the acceptable walking distances from an individual's house are presented and analysed. Walking distances from a parked car or from a station are of some interest but are less compatible with the research on walkability and the integration with the LifeLines data. This is because these analyses were conducted from the respondent's home to a particular utility.

The distance to walk to public transport depends on various factors, both internal and external to urban areas. For example, an Australian study shows that the degree of transport strongly influences the acceptable walking distance. When public transport is faster, has a higher frequency and covers a longer distance, people are more willing to walk. For a public transport stop, a person is prepared to walk approximately 450 metres, whereas to a stop for high-quality public transport that same person considers 800 metres an acceptable walking distance (Daniels & Mulley, 2011). NS (Dutch Railways) conducted a similar study for the Netherlands, which revealed that about half of the passengers are prepared to walk 1.3 kilometres to a station (CROW, 2021). Table 4.1 shows the acceptable walking distances in metres from the most recent publication (2021) by CROW.

Relocation	**Acceptable walking distance (in meters)**
From home to parked car	100 – 200
From home to bus stop (local)	200 – 500
From home to bus stop (interlocal)	250 – 900
From home to supermarket	450 – 1000
From home to city center / shopping area	500 – 1500
From home to work	250 – 1000
From home to school	250 – 900
From home to café / bar / restaurant	500 – 1000

Table 4.1. Acceptable walking distances (CROW, 2021)

According to the questionnaires from CROW, there are a number of spheres of influence that determine the acceptance of walking distances.

The most important spheres of influence according to CROW are:

- The attractiveness of the walking route;
- The directness of a walking route (and the possible presence of barriers);
- Social safety and road safety of the walking route;
- The parking regulation and pricing;
- The physical condition of a pedestrian;
- The motive of the visit (recreational/utilitarian);
- The waiting time (for example at a crossing);
- The duration of stay/visit location.

In addition to influences from the built environment, socio-demographic and geographical factors also influence the walking behaviour of individuals. To understand the Dutch walking culture, it is important to understand these differences.

Socio-demographic factors

Harms (2007), Heinen (2011), Bonham & Wilson (2012), Garrard et al. (2012), Scheepers et al. (2013), CROW (2014) and Harms et al. (2014) all describe six demographic characteristics that mainly influence the use of active modes. These are; age, gender, income, educational level, ethnic background and household composition.

Age

Between age groups, there is a clear difference in the extent to which people walk. Figure 4.1 illustrates the proportion of walking per age group. The groups that walk most are children under 12 and older people aged 65 and over (de Haas & Hamersma, 2019). This is mainly due to limited options, for example, they do not (or no longer) have a driving licence. From the age of 12, the degree of transport shifts mainly to cycling, the group between 12 and 17 years cycle the most of all age groups with a share of 41% (Schaap et al., 2015). The group up to 35 years of age walks relatively less. However, this group is responsible for 44% of all walking movements. The purpose of walking movements also plays a role with age. For example, the group between 50 and 75 years of age walks relatively the furthest, this is mainly due to walking as a leisure activity (de Haas & Hamersma, 2019). Although the elderly walk relatively the greatest distance, there is a positive correlation between an older age and a sedentary lifestyle. Older people walk less than younger people. However, when they do walk, it is mainly for recreational purposes and they also walk a longer distance. (O'Donoghue, Perchoux, & Mensah, 2016).

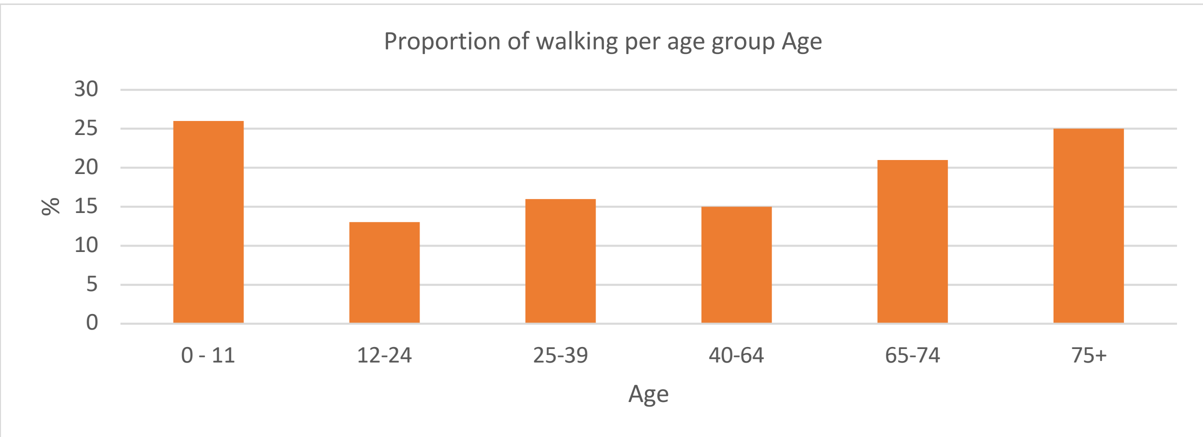


Figure 4.1. Walking proportions per age group (de Haas & Hamersma, 2019).

Gender
When a comparison is made between different genders, it appears that women walk more often than men, 60% versus 40%. One of the reasons why women walk more often is because they carry out activities that can be accomplished more often by walking, such as shopping. Of all trips made by women, about 19% are made on foot compared to 16% for men (de Haas & Hamersma, 2019). In addition, men aged 25 and over make more use of the car than women, this has various causes. For example, Schaap et al. (2013) state that women more often pick up children from school, walking or by bicycle, where men more often do so by car. In addition, women more often work part-time and closer to home (Schaap et al, 2015).

Income
Income also appears to be an important determinant of walking movements as can eb seen in figure 4.2. It can be stated that the lower the income, the more walking movements there are, while people with a high income make more use of other modes of transport and less use of 'active modes' (Schaap et al., 2015). This is partly due to the type of work carried out and the associated work environment that determines the level of physical activity and sedentary behaviour (Stappers, Schipperijn, & Kremers, 2020).

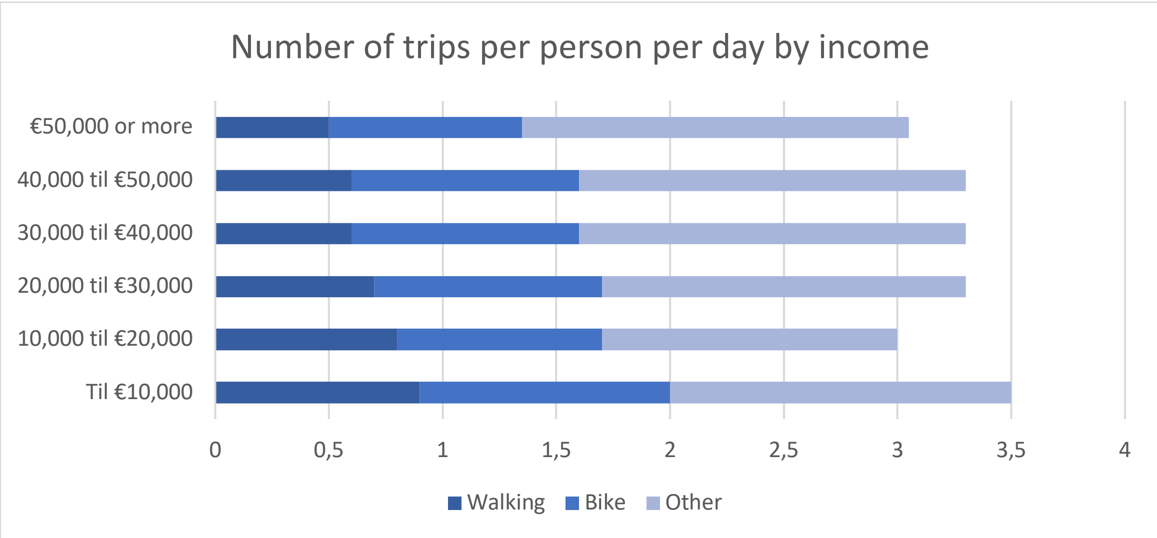


Figure 4.2. Walking trips per income group (Schaap et al., 2015)

Educational level
The level of education also affects the proportion of walking trips. The lower educated walk relatively more often than the higher educated, mainly because the lower educated generally have shorter commuting distances. This group also consists more often of young and elderly people, and it was already known that they walk more often than other population groups (de Haas & Hamersma, 2019). However, this relationship cannot be seen everywhere in the Netherlands. Especially in the big cities, a clear distinction in walking movements and education can be seen, where in the rest of the Netherlands this is less visible (Schaap et al., 2015).

Ethnic background
When active modes are compared to ethnic background, a clear difference can be seen. Individuals with a (non-)Western background walk more often than individuals with a Dutch background, 24% versus 17% of all trips. In contrast, individuals with a Dutch background cycle more often than individuals with a non-Western or Western background (de Haas & Hamersma, 2019). There are various reasons for this; for instance, immigrant children are more likely to walk to primary school because their parents do not cycle much and sometimes even consider it dangerous. It should be noted, however, that this differs substantially per city; the city of Rotterdam, for instance, has a large

difference between the number of walking and cycling trips, whereas this difference is less pronounced in rural areas or other (medium-sized) cities (Schaap et al., 2015).

Household composition
The household composition also plays a role in the proportion of cycling and walking. It makes a difference whether a couple has children or not. A couple with children on average makes about 3.2 trips per person per day compared to 2.6 trips per person per day for couples without children. This means that couples with children are about 20% more mobile than couples without children. However, differences can also be seen between single-person households and households with children. Households with children make less use of active transport options than single-person households, they use the car more often.

Geographic determinants
In addition to differences in socio-demographic characteristics, there are also major differences in terms of geography. For example, the proportion of walking differs greatly between urban and non-urban areas. It appears that residents in highly urbanised areas walk relatively much and also use public transport to a greater extent. This meets the expectations, in highly urbanised areas the public transport network is usually better aligned and there are more alternatives than in less urban or rural areas. This is especially visible when comparing the travel movements of young adults. Young adults in highly urbanised areas use public transport more often and walk and cycle more than the same group in non-urban areas (Schaap et al., 2015). A large group responsible for this purpose are students in large and medium-sized cities. They rarely own a car and can mostly use public transport for free. However, between cities large differences can still be seen when it relates to walking movements. These differences are mainly due to the different composition of the population.

The Mobility Assessment (Mobileitsbeeld) 2014 examined how these differences between cities could be explained. They devised four factors that strongly influence the walking behaviour of inhabitants. Firstly, the socio-cultural composition of the population. People with a foreign background cycle less, but walk more often than people with a native background. A second factor is the socio-economic composition of cities. In cities with lower incomes and relatively high unemployment, such as Heerlen, Arnhem and Tilburg, there is considerably more walking than in cities with a higher average income, such as Haarlem and Leiden. Thirdly, the demographic composition is important. Cities with a relatively young population, 25 to 45 years old, walk and cycle relatively more than regions with a relatively older population. This too can be partly explained by the presence or absence of students. Finally, the type of household is also a determinant for differences between urban areas. In areas with many single-person households and often students (Groningen, Nijmegen, Maastricht), there is relatively more walking and cycling than in areas with relatively more families (Apeldoorn, Dordrecht, Haarlem) (Schaap et al., 2015).

Urban and Rural
Differences between rural and urban communities can be seen in the Netherlands as a whole; for instance, people in the Randstad and in the south of Limburg walk relatively more than in the eastern part of the Netherlands (de Haas & Hamersma, 2019). Furthermore, almost all urban areas have a larger proportion of walking movements than rural areas, except for the south of Limburg, which probably has to do with a large proportion of walking for leisure. The low proportion of walking movements in the northeast of the Netherlands is remarkable. This could be explained by the relatively greater distance between urban areas and the large presence of rural areas.

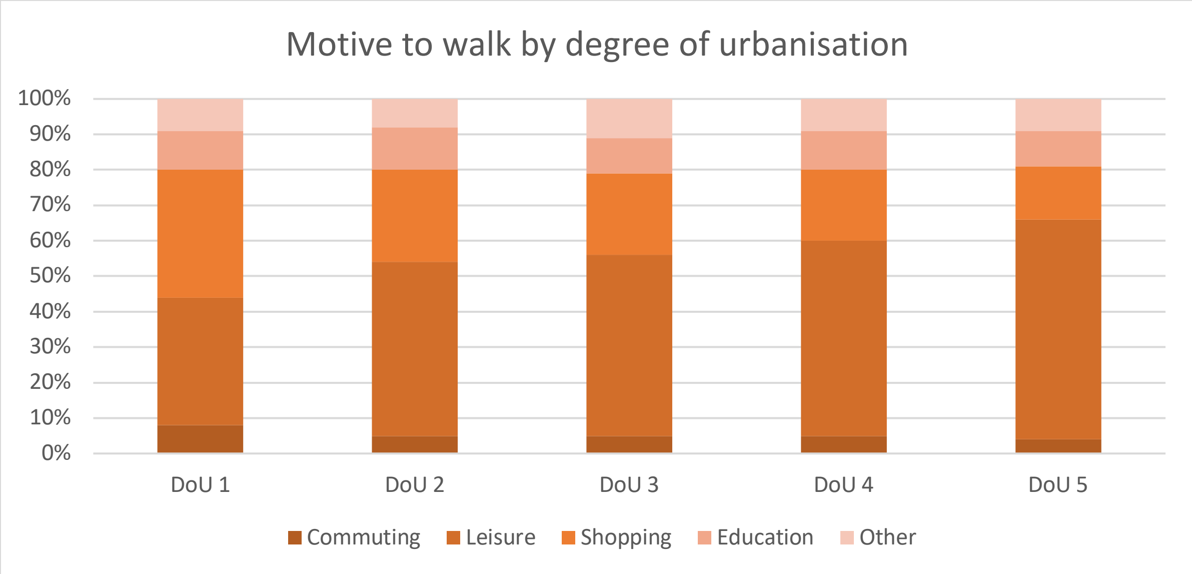


Figure 4.3. Motives for walking subdivided by DoU (de Haas & Hamersma, 2019)

Figure 4.3 displays clear differences in walking purposes between urban and rural areas. As urbanisation decreases, people walk more for leisure or to shop, compared to urban areas where people walk more for education; here again, students probably play a major role. In the Netherlands, there is no significant amount of walking for commuting; this is most likely due to the extensive use of bicycles.

The variables mentioned above largely explain the proportion of walking or cycling in different geographical areas. However, differences within cities themselves should also be noted. For instance, there are large differences between different neighbourhoods in a certain area that are determined in one neighbourhood by the difference in demographic compositions and in other neighbourhoods by the difference in cultural compositions. A background characteristic such as religion also shows differences in the proportion of people who walk. For example, it has been shown that a high proportion of Protestants has a positive effect on the extent of cycling but a negative effect on the extent of walking. On the other hand, in areas where Catholics predominate, there is a higher rate of walking and a lower rate of cycling. However, this again can vary greatly within a given urban area. It is therefore important to consider that there are differences within urban areas. In most analyses, these differences are taken together, and a statement is made about the entire area, so this may differ from reality.

Conclusion

The Dutch walking culture is not easy to determine by definition. Several demographic factors show differences in walking distances and there are multiple explanations for these differences. In addition, geographical areas differ from each other; for example, individuals in urban areas walk for different purposes than individuals in rural areas, and a difference in distance can also be observed. In urban areas, people walk relatively more than people in rural areas, reasons for this can be found in a denser population density and multiple utilities within walking distance. In rural areas, this is less common and people are more likely to use their cars. In addition, the public transport network in rural areas is less dense than in urban areas. A trend that seems to be continuing because areas where the public transport network is not profitable enough are seeing the threat of public transport disappearing or the frequency being reduced (Delis, 2017). This is partly due to the acceptable walking distance that people are prepared to take in order to reach a certain final destination; these are often too high in rural areas and so people tend to choose other alternatives.

It can be concluded that most acceptable walking distances to reach a certain utility lie between 200 metres and 1000 metres. In the publication by CROW, walking as a leisure activity is not included, so this distance is generally longer and distances up to two kilometres are quite common. That is one of the reasons that GECCO has drawn up several WI with different buffer sizes in their research. In this results section, the results of the WI with a buffer of 500 metres and with a buffer of 1650 metres are therefore examined for the PC6-approach and the RP-approach.

4.2 500 metres approaches

First, the results of the 500-metre analyses are identified, analysed and explained. These are the analyses in which each component is calculated with a Euclidian buffer range of 500 metres of a certain location. If all seven components are within 500 metres of a location, then that location will receive a high walkability score. In practice, this mainly applies to small sized, urban areas, where many utilities are located together. Also, in urban areas most of the seven components are relatively more present than in rural areas. For example, population density, land use mix, sidewalk presence, street connectivity and transportation density are more present in urban areas.

A total of 36,366 PC6 areas were compared in this study. These are all PC6 areas where LifeLines participants live. To make comparisons, these PC6 areas were divided according to their degree of urbanisation as determined by the OAD by the CBS. This has shown that a relatively large number of PC6 areas have an OAD of an average of less than 500 addresses per km2; these are the 15.895 PC6 areas with a score of 5 (non-urban). These areas cover more than 43% of the total number of PC6 areas in this study (see Table 4.2). It appears that the larger the PC6 areas, the more PC6 there are. This may indicate that there are relatively few urban areas in the northern parts of the Netherlands.

DoU

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-99997	14	,0	,0	,0
	1	3531	9,7	9,7	9,7
	2	3714	10,2	10,2	20,0
	3	5864	16,1	16,1	36,1
	4	7354	20,2	20,2	56,3
	5	15895	43,7	43,7	100,0
	Total	36372	100,0	100,0	

Table 4.2. Frequency of PC6 areas per DoU

The analyses using the PC6- and RP-approaches were intended to provide a more accurate index, particularly for rural areas. This is because the GECCO approach was based on a centroid, the centre of a PC6 area. The PC6-approach is based on all grid values in the PC6 area, the RP-approach is based on the locations where people live. To conclude whether both approaches differ significantly or to what extent these approaches correlate with each other and with the centroid approach, the WI are compared in both ArcGIS and SPSS. To validate if the degree of urbanisation affects the difference in the WI per PC6 area, they are compared dependent on each other. For example, the areas with an OAD of 1 are compared with each other. Finally, the urban (1) and rural values (5) are compared to illustrate the applicability per PC6 area. It will be limited here to being able to demonstrate differences or similarities between the various approaches. This is because it is difficult to be able to show from the available data at the PC6 level which approach is better and therefore predicts the WI better. This research is therefore limited at this stage to demonstrating different approaches and how that relates to different WI and different degrees of urbanisation.

4.2.1 Centroid to PC6

In this study, three approaches to calculate a WI are examined and compared. The GECCO approach, the Centroid-approach, is compared with both the PC6- and the RP-approach to illustrate the differences. The Centroid-approach assumes the centre of a PC6 area, but unlike the PC6-approach, large parts of certain PC6 areas are not included in this calculation. This results in an inaccurate WI-score for the PC6 area and this problem increases as the PC6 area gets larger, particularly in rural PC6 areas.

To understand the differences between the Centroid- and PC6-approach, the extent to which both WI correlate is first examined. As most areas in the northern provinces are rural, a situation arises in which the large majority of PC6 areas receive a relatively low score based on the calculations with a 500-metre buffer. Only the urban areas in these provinces score higher on average on the WI. However, as Table 4.2 displays, less than 10% falls within a degree of urbanisation of 1, or a density of 2500 addresses per km/2. These city centres are less relevant because the PC6 areas there are relatively small, which implies that it does not significantly matter whether a Centroid-approach or a PC6-approach is used in such a PC6 area. Most PC6 areas in strongly urbanised areas are smaller than 500 metres, a point in the middle of this area or calculating all values in this same area will probably generate a less significant difference than in a large rural area.

Table 4.3 shows, on the degree of urbanisation, whether the score of a PC6 area has changed positively or negatively after performing the PC6 analysis. A DoU of 1 is strongly urban and a DoU of 5 is strongly rural. It appears that as the degree of urbanisation decreases, more PC6 areas receive a lower WI-score. In other words, when the PC6-approach is applied instead of the Centroid-approach, 85% of the PC6 areas with an urban degree of 3 receive a lower WI-score. With an urbanity of 1, about half of the PC6 areas have received a lower WI-score and half have received a higher score. However, with an urban degree of 4, no less than 90% of the PC6 areas have a lower score in the PC6-approach than in the Centroid-approach. With an urban degree of 5, that number is 72%, which could be due to the high number of PC6 areas (15,895) compared to the other urban degree levels. It appears that the difference between the two approaches increases as the degree of urbanisation decreases.

Centroid to PC6					
	DoU 1	**DoU 2**	**DoU 3**	**DoU 4**	**DoU 5**
Change positive +	1819	860	884	749	4399
Change negative -	1666	2794	4958	6587	11.461
% low	48%	76%	85%	90%	72%

Table 4.3 Change in WI-score between PC6- and Centroid-approach 500-metre buffer

To understand the correlation between the two approaches, a scatterplot is generated. This scatterplot shows on the Y-axis the value of a PC6 area based on the Centroid-approach. The X-axis shows the value of that same PC6 area but based on the PC6-approach (figure 4.4). If a value lies below the horizontal R2 line, that PC6 area has a higher outcome in the PC6-approach than in the Centroid-approach. If the value of the PC6 area is above the line, that area will have a lower score in the PC6-approach than in the Centroid-approach. These values are also compared based on the degree of urbanisation because a statement for all 36,366 PC6 areas will be distorted by the difference between urban and rural PC6 areas. The expectation prior to this analysis was therefore that as the degree of urbanisation decreases, the two approaches correlate less and less with each other. The correlation (R2) indicates the extent to which the Centroid-approach correlates with the PC6-approach. In other words, to what extent is there a statistical correlation between two quantities, this is expressed by the correlation coefficient. The presence of a correlation does not mean that there is a causal relationship between the two approaches. The correlation only indicates the extent to which the two variables are related. The value of a correlation coefficient can lie between -1 and 1, whereby a value of -1 indicates a negative relationship and a value of 1 indicates a positive relationship. A value of 0 indicates no correlation. A value between 0.5 and 0.7 is considered a moderate effect size. If the R2 is higher than 0.7, this is considered a strong effect. This high value means that the model explains almost all of the variability of the response data around the mean. The lower this value, the less the independent variable explains the variation in the dependent variable (Moore, Notz, & Flinger, 2013).

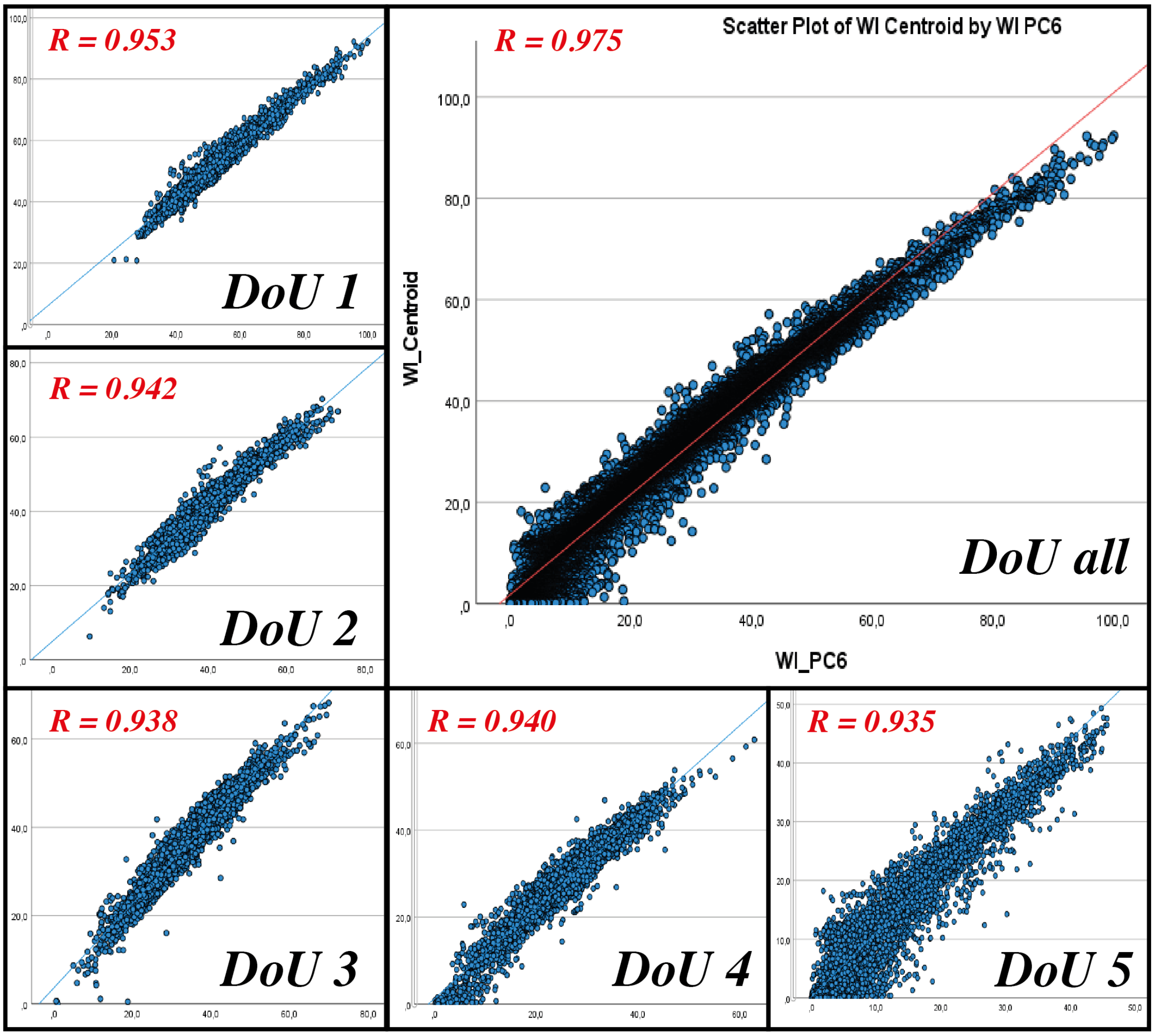


Figure 4.4. Scatterplots of Centroid- and PC6-approach 500-metre buffer

Figure 4.4 shows a scatterplot comparing the correlation between all PC6 areas in the Centroid-approach and the same PC6 areas in the PC6-approach. In this case, there is a positive correlation with a correlation coefficient of 0.975 and most of the values are close to each other with a few exceptions. This correlation is very high and it implies that most of the PC6-approach does not deviate greatly from the Centroid-approach. However, in this model there are PC6 areas in all degrees of urbanisation. To detect whether the differences are larger or smaller based on urbanisation, the same models are made per degree of urbanisation. The expectation is that as the degree of urbanisation decreases, so does the correlation coefficient. If so, the Centroid-approach is compared in the same way with the RP-approach to understand which of the two approaches generates a greater difference.

The figure below shows that there is little difference in the two approaches. All degrees of urbanisation have a correlation coefficient higher than 0.9, which indicates that in each degree of urbanisation a very strong correlation can be found between both variables. However, it is noticeable that the correlation is highest in urban areas and lowest in rural areas, although these differences are minimal. A degree of urbanisation of 1 results in a correlation coefficient of 0.953, while an urbanisation of 5 results in a correlation coefficient of 0.935. Furthermore, it can be seen that the WI-score range decreases as urbanity decreases.

4.2.2 Centroid to RP

In the second approach, the Residential Property approach, the values of the PC6 areas are determined by only including those values in the calculation that intersect with the area of a residential property. In contrast to the PC6-approach, large sections of the PC6 areas are therefore not included in the creation of these WI values.

This approach seems more fair when determining a WI. A walkability index that consists of components such as population density, sidewalk presence, land use mix, or transportation density is based on places where people live. A large piece of arable land or forest has a lower chance of a high WI-score because most WI components do not occur here, there are no sidewalks or land use mix in a forest area. It is therefore not logical to include areas where no people live in the calculation of a WI. A PC6 area with a surface area of 10,000 km2 that only includes two farms will almost always receive a low WI-score as the areas without human activities are included in the calculation of the WI-score for that PC6 area. The RP-approach tries to minimize this margin of error by only looking at the values under residential properties. Also in this approach, the WI-score can be very low, especially when someone lives in a remote area with less human activity. However, the WI-score for this PC6 area is determined by the locations where people live and is not or less influenced by values on the other side of the PC6 area as is the case in the PC6-approach, where all values in that PC6 area are included in the calculation.

As in the previous approach, the RP-approach is contrasted with the Centroid-approach to show whether there are predominantly positive or negative changes in the WI-scores of the PC6 areas. Table 4.4 shows that, as in the PC6-approach, the RP-approach results in relatively more negative WI-scores than positive scores. Also, in the RP-approach, more PC6 areas receive lower scores when urbanity decreases. The PC6 areas with an urbanity of 1 change the least, here slightly more than half receive a lower score. It is noticeable that a DuO of 5 generates almost the same number of lower PC6 areas in the RP-approach as in the PC6-approach, both 72%.

Centroid to RP					
	DoU 1	**DoU 2**	**DoU 3**	**DoU 4**	**DoU 5**
Change positive +	1434	696	838	937	4445
Change negative -	2053	2978	5003	6398	11.410
% low	59%	81%	86%	87%	72%

Table 4.4 Change in WI-score between RP- and Centroid-approach 500-metre buffer

To determine the extent to which the RP-approach correlates with the Centroid-approach, these two approaches are also contrasted in a scatterplot. Here it also applies that a high correlation means a high coherence. Figure 4.5 shows the correlation between all 36,366 PC6 areas in the Centroid-approach and the RP-approach. With a correlation coefficient of 0.936, the correlation is high, meaning that the model explains almost all the variability of the response data around the mean. A notable feature of this matrix is that mainly the PC6 areas that received a WI-score of zero in the Centroid-approach have increased in score in the RP-approach. However, as Figure X shows, most PC6 areas in the RP-approach gained a lower score than in the Centroid-approach. The Centroid-approach has many WI-scores of zero compared to the PC6-approach and RP-approach. The fact that fewer zero scores can be found in the RP-approach is because this approach is based on dwellings. Residential properties where people live almost always have higher scores in one of the components than when the centre of a PC6 area is assumed.

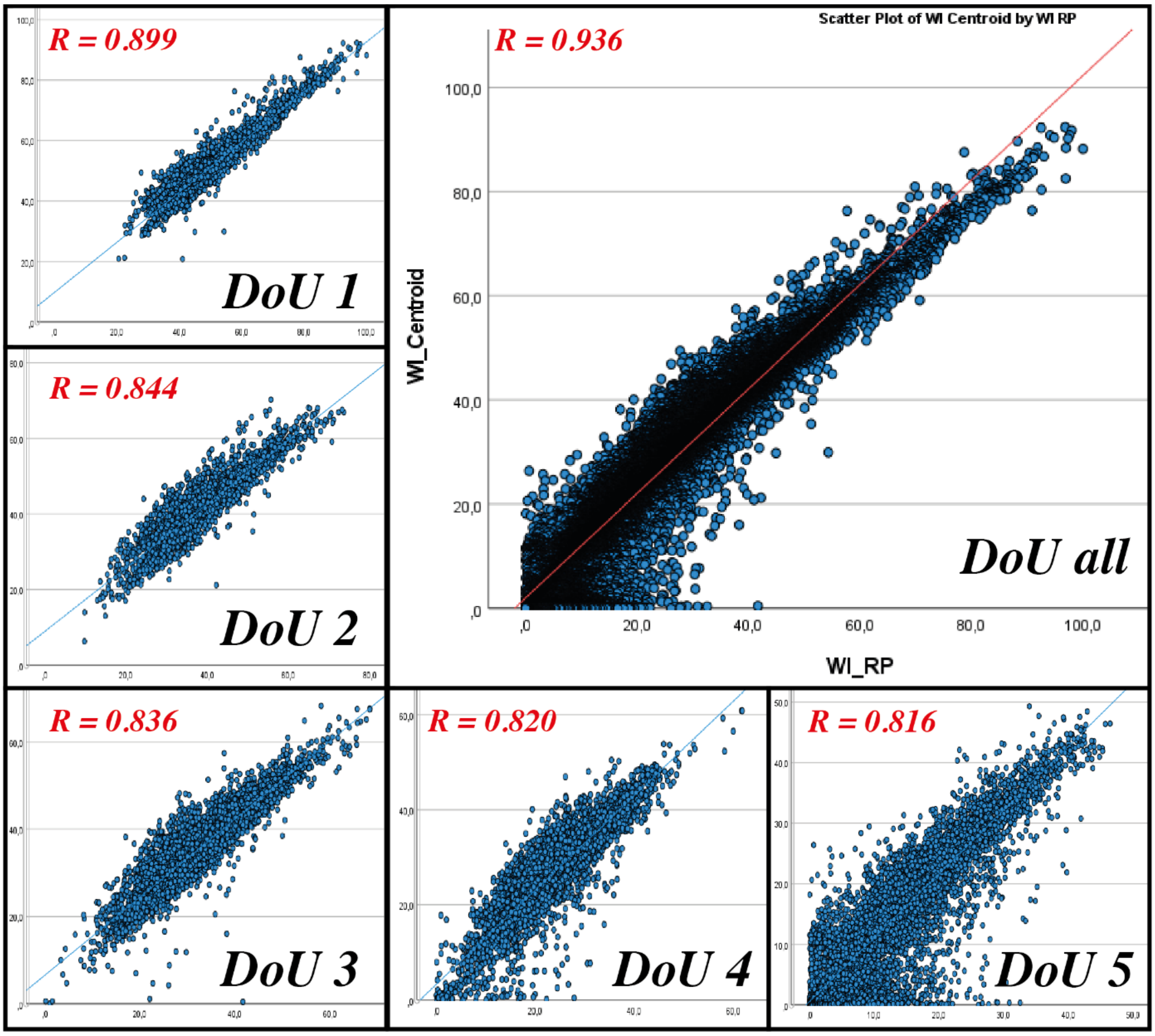


Figure 4.5. Scatterplots of Centroid- and RP-approach 500-metre buffer

In rural areas, the Centroid-approach therefore relies mainly on chance and to a lesser extent on the correctness of the number of components in that area. In order to properly demonstrate the distinction between rural and urban areas, different scatterplots based on urbanisation are also drawn for this approach.

Figure 4.5 shows the scatterplots between the Centroid-approach and the RP-approach. In terms of the difference between rural PC6 areas and urban PC6 areas, a greater difference can be seen than in the previous approach. By a degree of urbanisation of 1, the correlation coefficient is 0.899, and as urbanisation decreases, the correlation coefficient also decreases. With an urban degree of 5, the correlation coefficient is 0.816, although this means that there is still a high degree of mutual correlation. However, these scatterplots do show that the correlation between the two approaches decreases as the degree of urbanisation decreases, which could be an indication that the RP-approach can generate a more accurate score for the WI. For example, the scatterplot with an urbanity of 5 shows that more outliers can be found than with a higher degree of urbanity. Many PC6 areas that receive a score of 0 in the Centroid-approach, receive a higher score in the RP-approach up to more than 30. However, the opposite also happens where a score of 0 in the RP-approach leads to a higher score in the Centroid-approach.

It is difficult to make a statement based on a correlation coefficient alone. A high or low correlation does not necessarily imply that one of the two approaches works better in terms of calculating a WI-score. It can be concluded that PC6 areas receive different scores based on the chosen approach and it can be stated that there is more variation in RP-approach than in PC6-approach. As expected, this variation increases with decreasing urbanity and increasing size of the PC6 areas. In the next section, the PC6-approach is compared with the RP-approach to identify the extent to which these two methods correlate with each other and what differences can be found.

4.2.3 PC6 to RP

Both the PC6-approach and the RP-approach show a high degree of correlation with the Centroid-approach even though almost all PC6 values have changed. In both approaches, this has resulted in relatively lower WI-scores but, as the scatterplots show, the change in WI-scores for most of the PC6 areas is marginal. The PC6 areas that changed most after applying one of the two approaches are compared in the next section. This section illustrates the extent to which the PC6-approach correlates with the RP-approach and the extent to which the degree of urbanisation affects this correlation.

PC6 to RP					
	DoU 1	**DoU 2**	**DoU 3**	**DoU 4**	**DoU 5**
Change positive +	805	964	1741	2551	7040
Change negative -	2668	2674	4075	4751	8763
% low	76%	73%	70%	65%	55%

Table 4.5 Change in WI-score between PC6- and RP-approach 500-metre buffer

Comparing the RP-approach with the PC6-approach, table 4.5 shows that most of the PC6 areas receive a lower score in the RP-approach. Mainly in the highly urbanised areas, 76% of the PC6 areas receive a lower score in the RP-approach than in the PC6-approach. This percentage decreases as the degree of urbanisation decreases; with a DoU of 5, only 55% of the PC6 areas obtain a lower score. This is remarkable, as the expectation was that the RP-approach would work more effective for rural areas, as there are relatively fewer residential properties in the large rural areas than in the urban PC6 areas. Figure X shows that this is not true, as slightly more than half of the strongly rural PC6 areas receive a lower score in the RP-approach. When examining the mutual correlation, it appears from table 4.6 that both approaches correlate strongly. As urbanisation decreases, the correlation also decreases slightly, but the mutual correlation is still high.

PC6 to RP	**DoU 1**	**DoU 2**	**DoU 3**	**DoU 4**	**DoU 5**
Correlation coefficient	0.942	0.902	0.903	0.891	0.887

Table 4.6. Correlation coefficient between PC6- and RP-approach 500-metre buffer

To determine which approach better predicts the WI score, the PC6 areas with the greatest differences between the two approaches will be compared. Both the PC6-approach and the RP-approach generate more lower scores for the PC6 areas than the Centroid-approach does. When comparing the two adjusted approaches, the RP-approach generates more lower scores than the PC6-approach as shown in figure 4.5. Using the ten PC6 areas with the largest differences, it can be visually determined which approach is more accurate. Table X shows the PC6 areas with the largest positive and negative (in italic) difference between the PC6-approach and the RP-approach. Those PC6 areas received a relatively high or low WI-score in the RP-approach compared to the PC6-approach.

PC6	WI_Centroid	WI_PC6	WI_RP	Diff. C_PC6	Diff. C_RP	Diff. PC6_RP
9244GC	0.38	5.03	30.88	4.65	30.50	25.85
9713AB	29.94	31.14	54.32	1.20	24.38	23.18
7908AD	0.47	18.87	41.70	18.41	41.23	22.83
8881BD	0.00	3.33	26.09	3.33	26.09	22.76
7823BG	35.42	30.18	51.19	-5.24	15.77	21.01
7855TD	*26.42*	*29.65*	*0.66*	*3.23*	*-25.76*	*-28.99*
7846TD	*22.93*	*28.57*	*3.22*	*5.64*	*-19.71*	*-25.35*
7871TE	*20.03*	*25.57*	*2.13*	*5.54*	*-17.90*	*-23.44*
7814RT	*24.84*	*31.25*	*8.08*	*6.41*	*-16.76*	*-23.17*
8427RH	*24.64*	*27.63*	*4.97*	*2.99*	*-19.67*	*-22.66*

Table 4.7. Ten biggest differences, positive and negative, between approaches with a 500-metre buffer

Table 4.6 shows that both approaches correlate strongly with each other, indicating that there are almost no extreme differences over the entire data set. Table 4.7 shows the PC6 areas in which the RP-approach found a relatively higher score compared to the Centroid-approach. However, because both approaches do not differ substantially from each other, this is a small proportion of all PC6 areas. Of the 36,225 PC6 areas, 32,509 PC6 areas in the RP-approach have a WI-score that is at most 5 lower or 5 higher than in the PC6-approach. Just over 10% of all RP PC6 areas have a WI-score that differs by more than 5 from the PC6-approach. In terms of the PC6 areas with a difference of up to 2.5 positive and up to 2.5 negative, this results in 28,796 PC6 areas. This means that almost 80% of all PC6 areas are within a difference of a score of 5 (2.5 negative and 2.5 positive). All differences can be seen in figure 4.6.

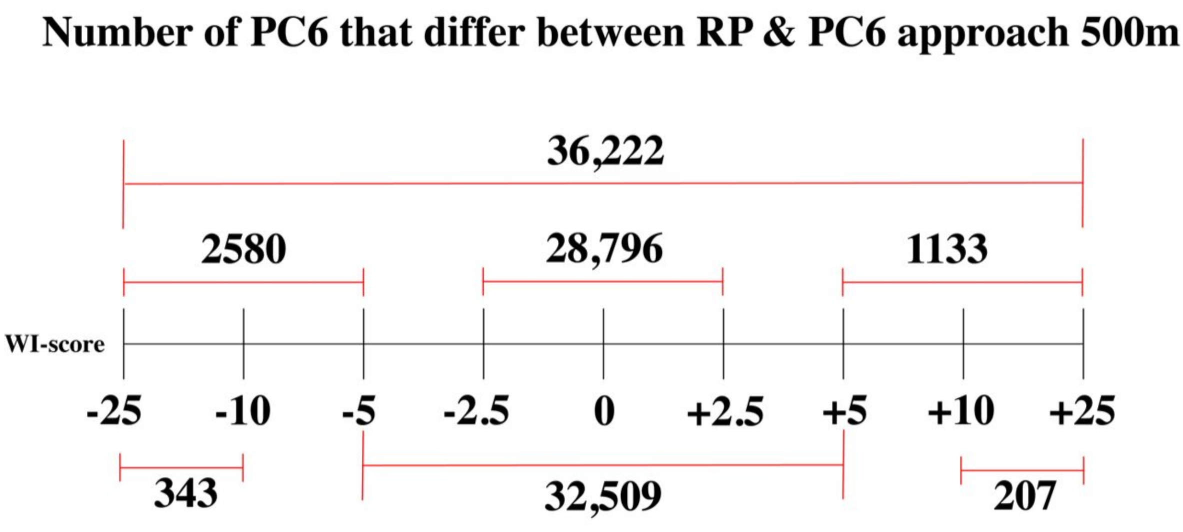


Figure 4.6. Differences in WI-scores RP- and PC6-approach 500-metre buffer

Between the two approaches, no great difference in WI-score can be observed for most PC6 areas. It is therefore mainly the rural and therefore the PC6 areas with a large surface area, that display the largest differences. These large differences in the rural PC6 areas can be well explained by means of the RP-approach;

4.2.3.1 Visual examples 500-metre buffer

Table 4.7 shows the biggest differences between the PC6- and RP-approach. Number one, postal code 9244GC, receives a low score (0.38 and 5.03 respectively) in both the Centroid-approach and the PC6-approach, while this same PC6 area receives a WI-score of 30.88 in the RP-approach. Figure 4.7 shows how this difference arises, the blue circles represent the buffer of 500 metres around each residence, this is the buffer that is used when calculating the 500-metre WI-score. In the Centroid-approach, this PC6 area, which consists of several areas, is given a value in the middle of these areas. In this case, there is virtually no WI-component that lies below the centroid, hence the score of 0.38. The PC6-approach calculates the score based on all the values in the PC6 area and in this case only the north-east gets a higher score, which is a relatively small area compared to the whole PC6 area, hence the score of 5.03. The RP-approach only calculates the scores under the red-coloured houses in the north and north-east. These are almost all in an area where there are also components within a radius of 500 metres; those areas with almost no components are not included in this calculation. Therefore, the RP-approach is more valid, especially since the WI-score says something about the people who live there and areas where nobody lives are therefore not relevant. The same applies to the PC6 area 8881BD (figure 4.8) which contains only residential properties in the north and none in the rest of the PC6 area.

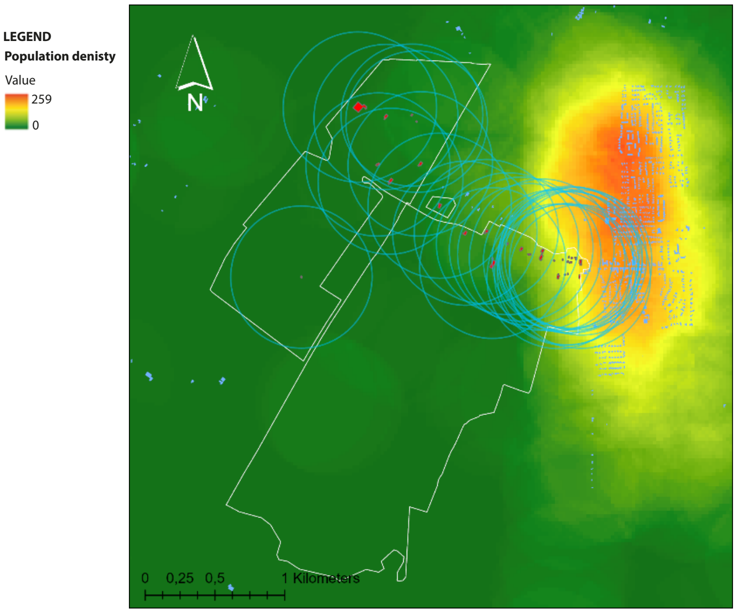


Figure 4.7. PC6 area 9244GC with 500-metre buffers

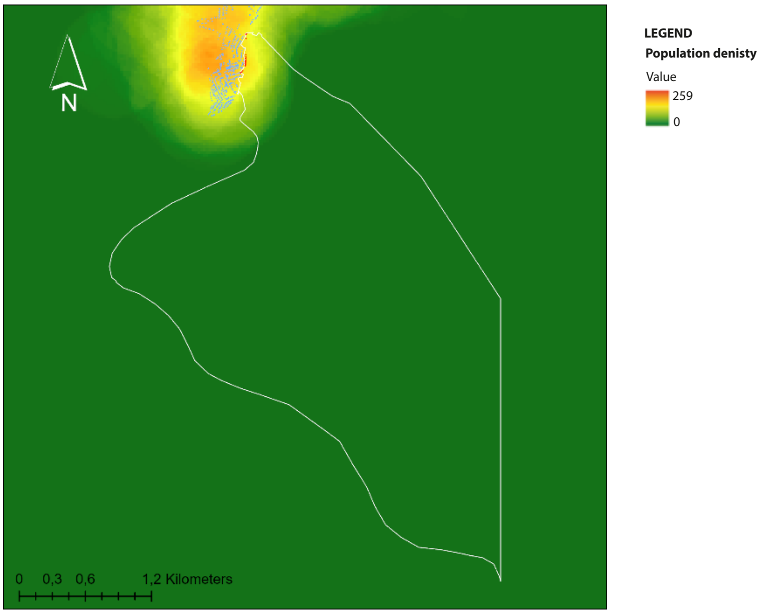


Figure 4.8. PC6 area 8881BD 500-metre buffer

A second cause of difference between the two approaches arises in PC6 areas that consist of several non-contiguous areas. The centroid is then far from accurate as can be seen in figure 4.9, the centroid (in green) is in the middle of the large PC6 area while the residential properties lie in the west of that PC6 area and in an unconnected PC6 area further to the west. The PC6-approach is also often inaccurate since these are divided areas in which people live in one area and few or none in the other PC6 area. However, in the PC6-approach, those areas with a very low population density are still included in the calculation, which results in inaccuracy. This is evident, for example, in figure 4.10. This PC6 area consists of three separate areas although only in one of the three areas people live. The elongated area in the west determines a significant part of the score due to its surface, yet no people live in this area. Therefore, the RP-approach is more accurate in this instance as well, as it only assumes areas where people live, in this example, the small PC6 area in the east.

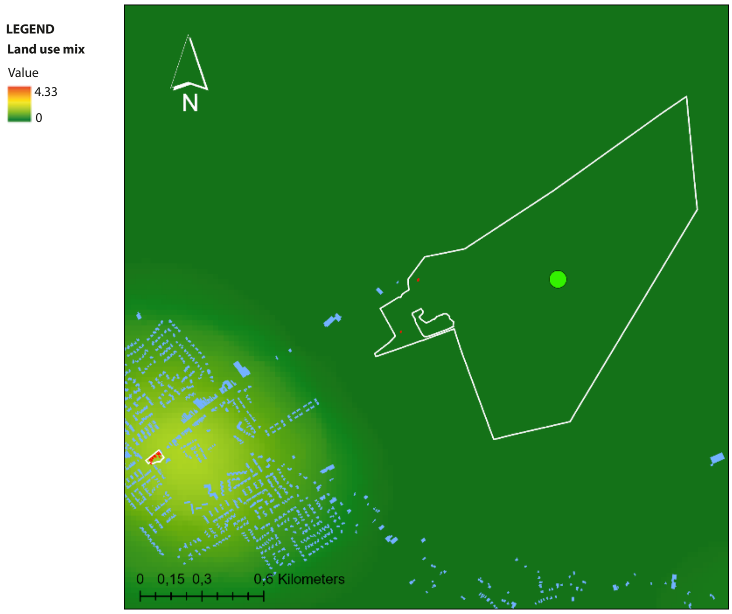


Figure 4.9. *PC6 area with centroid* 500-metre buffer

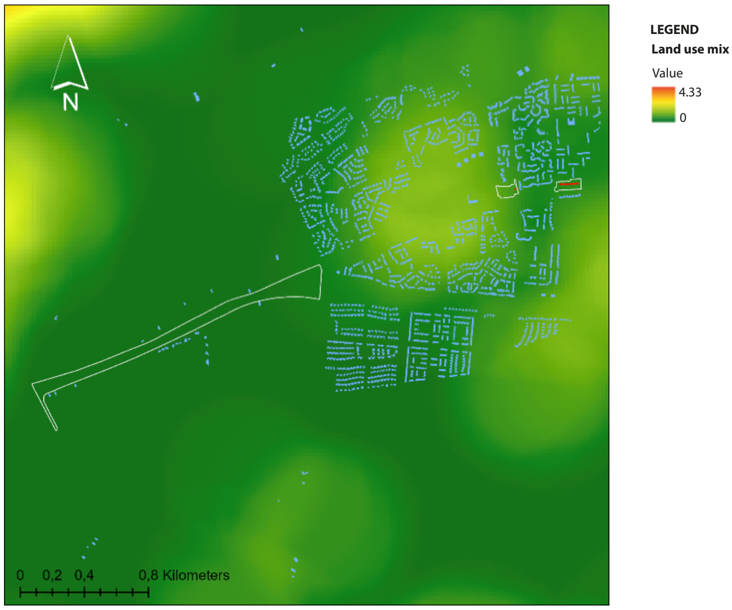


Figure 4.10. PC6 area non-contiguous 500-metre buffer

In addition, it may occur that in the RP-approach the PC6 areas receive a lower score, PC6 area 7855TD (figure 4.11) and 7814RT (figure 4.12) are examples of this phenomenon. This margin of error occurs mainly due to the 'green space density' component. Especially in rural areas, a large part of the PC6 area may consist of green space while there is no other component in that PC6 area. PC6 area 7855TD receives a high score in the Centroid-approach since the centroid lies within the green space component. The same applies to the PC6-approach, half of this PC6 area receives a high score for green space. The average score will therefore be higher compared to the RP-approach. Again, the RP-approach is more accurate, as the four residences in this area have no WI-component within a radius of 500 metres. The fact they obtain a high score because one of these components is present elsewhere in the PC6 area does not imply that this is also correct for these properties.

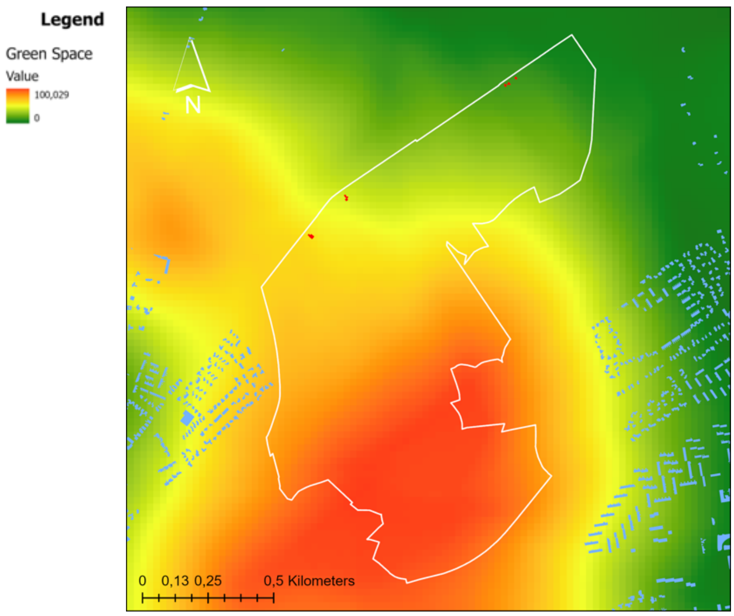


Figure 4.11. PC6 area 7855TD 500-metre buffer

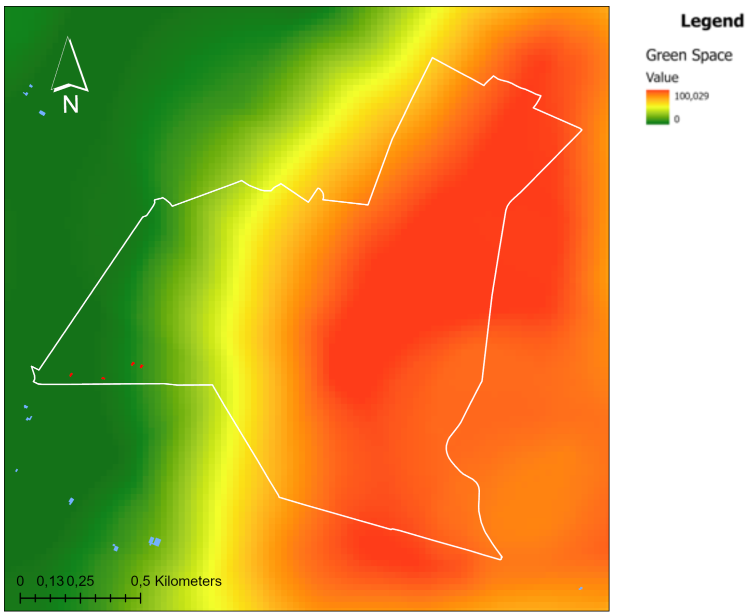


Figure 4.12. PC6 area 7814RT 500-metre buffer

Conclusion

Despite most PC6 areas not differing significantly in score when the PC6-approach or RP-approach is used, there are a number of large, rural areas where this does apply. The more urbanity decreases, the more differences between the approaches arise. It is therefore particularly common for PC6 areas with few residential properties and a large surface area to be significantly biased in the Centroid- and PC6-approaches. The RP-approach minimises this margin of error for these PC6 areas. It can be concluded that the PC6-approach and the RP-approach show similar results when compared to the Centroid-approach as both approaches correlate strongly, which indicates small differences in the WI-scores. At a buffer of 500 metres, the RP-approach is no worse than the other approaches and in an area such as the Northern Netherlands, containing many large rural PC6 areas, the RP-approach operates with the smallest margin of error. The majority of PC6 areas can be calculated using both the PC6-approach and the RP-approach, as both explain the WI-score more accurately than the Centroid-approach. However, to achieve an accurate and comparable result for all PC6 areas, the RP-approach results in the smallest margin of error.

4.3 1650-metres approaches

In addition to the 500-metre approaches, in which a buffer of 500 metres is applied, this study also calculates a WI with a buffer of 1650 metres. A residence in the 1650-metre approach will often get a different WI-score compared to the 500-metre approach. In the 1650-metre approach, more components could fall within the buffer than in the 500-metre buffer. The same applies the other way around; within the 1650-metre buffer, more areas could contain no single component, resulting in a lower average value. This study does not examine whether the 500-metre approach functions better than the 1650-metre approach, or vice versa, as a WI. The purpose of examining both approaches is to determine whether the PC6 and RP approaches show similar results regardless the buffer size.

The GIS analyses that were performed to convert the values per residential property to PC6 area required much computing power. The hardware available during this research could not perform all analyses at once, therefore for some components the areas were split up and these analyses were performed separately, then these separate analyses were merged again. However, by performing these separate analyses, some data went lost. This resulted in a loss of approximately 7,000 PC6 areas. Therefore, in the 1650-metre analyses, 28,539 PC6 areas are included instead of the original 36,366. The PC6 areas that are not included are spread across all levels of urbanisation.

4.3.1 Centroid to PC6

Comparing the Centroid-approach with the PC6-approach, it is remarkable that almost every PC6 area received a higher WI-score in the PC6-approach. This is in sharp contrast to the 500-metre analyses that mainly resulted in lower WI-scores after performing the PC6 analysis. In total, only two percent of all PC6 areas received a lower score in the PC6-approach, 98% of the PC6 areas scored higher. Based on these results, it appears that within a buffer of 1650 metres, considerably more components are included in the calculation compared to the Centroid-approach. This can be explained for the strongly urbanised areas; most urban areas are smaller than the 1650-metre buffer. It therefore appears that calculating a score in an urban PC6 area almost always results in a higher WI-score compared to the Centroid-approach. However, the same applies to rural areas, which can be explained by the fact that values outside the PC6 area, but within a radius of 1650 metres, are included in the calculated score for that area. Table 4.8 shows the differences between the Centroid- and PC6 approach with a 1650-metre buffer.

Centroid to PC6					
	DoU 1	**DoU 2**	**DoU 3**	**DoU 4**	**DoU 5**
Change positive +	3289	2793	4692	5745	11.424
Change negative -	0	0	0	7	575
% low	0%	0%	0%	0.1%	4.8%

Table 4.8. Change in WI-score between Centroid- and PC6-approach 1650-metre buffer

4.3.2 Centroid to RP

Similar to the PC6-approach, almost all WI-scores in the RP-approach result in higher scores compared to the Centroid-approach as can be seen in table 4.9. The same explanation as in the PC6-approach can be given, namely that a buffer of 1650 includes more values, also outside the PC6 area, in the calculation of the WI-score. Therefore, 2.7% of the PC6 areas receive a lower score in the RP-approach, a minimal difference with the PC6-approach.

Centroid to RP					
	DoU 1	**DoU 2**	**DoU 3**	**DoU 4**	**DoU 5**
Change positive +	3289	2791	4692	5706	11.285
Change negative -	0	2	0	46	714
% low	0%	0.1%	0%	0.8%	5.9%

Table 4.9. Change in WI-score between Centroid- and RP-approach 1650-metre buffer

4.3.3 PC6 to RP

The aforementioned tables have shown that both approaches generate almost the same results when compared to the Centroid-approach, almost all WI-scores being higher. When the PC6-approach is compared to the RP-approach (table 4.10), these differences are less extreme. However, it is noticeable that the RP-approach generates more higher scores than the PC6-approach in every degree of urbanisation, except by a DoU of 1. These differences become larger as urbanisation decreases, which contrasts with the 500-metre approaches where the difference decreased as urbanisation decreased.

PC6 to RP					
	DoU 1	**DoU 2**	**DoU 3**	**DoU 4**	**DoU 5**
Change positive +	1580	1597	3194	4418	8957
Change negative -	1709	1196	1498	1334	3042
% low	%	%	%	%	%

Table 4.10. Change in WI-score between PC6- and RP-approach 1650-metre buffer

To ascertain the extent of these differences in score, the correlation for both approaches is calculated per level of urbanisation. Table 4.11 shows that both approaches correlate strongly with each other; as urbanisation decreases, the correlation also decreases, but minimal and the correlation remains above 0.9 for each level. This means that both approaches differ minimally with respect to all PC6 areas, it is therefore difficult to make a statement about the effectiveness of both approaches based on their correlation.

PC6 to RP	**DoU 1**	**DoU 2**	**DoU 3**	**DoU 4**	**DoU 5**
Correlation coefficient	0.993	0.984	0.989	0.973	0.931

Table 4.11. Correlation coefficient between PC6- and RP-approach 1650-metre buffer

To enable a comparison between the two approaches and determine if the RP-approach predicts the WI-score better than the PC6-approach, table 4.12 shows the PC6 areas with the highest positive and negative differences. This table shows the WI-scores that were extremely higher or extremely lower in the RP-approach compared to the PC6-approach.

PC6	WI_Centroid	WI_PC6	WI_RP	Diff. C_PC6	Diff. C_RP	Diff. PC6_RP
9713AB	52.56	57.62	85.84	5.06	33.28	28.21
9163GV	0.00	0.98	28.06	0.98	28.06	27.08
9244GC	15.15	17.10	43.58	1.96	28.43	26.48
8862PK	0.06	9.65	32.65	9.59	32.59	23.00
8881BD	0.00	7.15	29.96	7.15	29.96	22.80
9746TJ	*30.10*	*33.81*	*9.71*	*3.70*	*-20.39*	*-24.09*
7814RT	*34.98*	*52.22*	*28.89*	*17.24*	*-6.09*	*-23.33*
9738TC	*14.27*	*24.44*	*8.21*	*10.17*	*-6.06*	*-16.23*
8765LR	*22.87*	*23.58*	*7.58*	*0.71*	*-15.30*	*16.00*
9523TH	*20.10*	*26.65*	*10.90*	*6.54*	*-9.21*	*-15.75*

Table 4.12. Ten biggest differences, positive and negative, between approaches with a 1650-metre buffer

As in the 500 m approach, no significant differences can be found between the PC6-approach and the RP-approach, it is evident that both approaches correlate strongly. Figure 4.13 illustrates that 26,887 PC6 areas have a WI-score in the RP-approach that is at most 2.5 lower or 2.5 higher than in the PC6-approach. There are only a few PC6 areas that receive an extremely higher or extremely lower score. These are fewer than in the 500-metre analyses, but this is partly because fewer PC6 areas are included in the 1650-metre approach. As the two approaches do not differ greatly, these extreme values are particularly interesting.

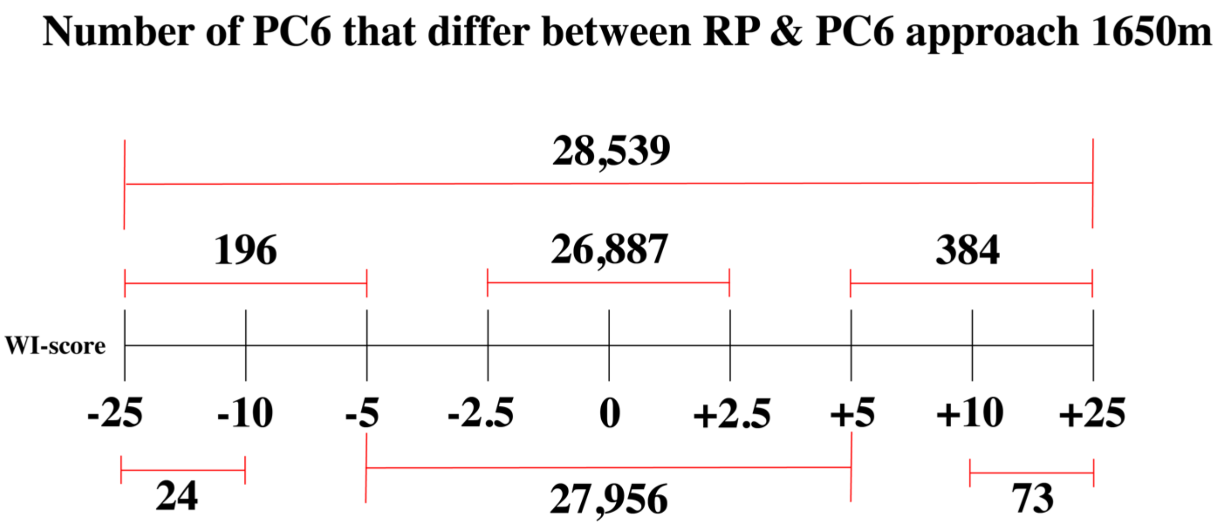


Figure 4.13. Differences in WI-scores RP- and PC6-approach 1650-metre buffer

4.3.3.1 Visual examples 1650-metre buffer

To demonstrate that these extreme values are indeed better explained in the RP-approach, the PC6 will be visually compared with each other. The PC6 areas in table 4.12 are visually illustrated in the examples below to show the extent to which the RP-approach is more accurate than the PC6-approach.

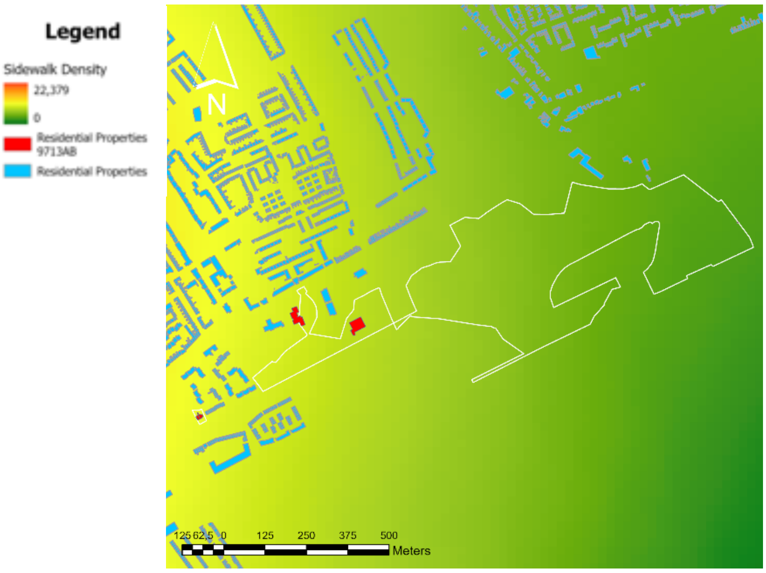


Figure 4.14. PC6 area 9713AB 1650-metre buffer

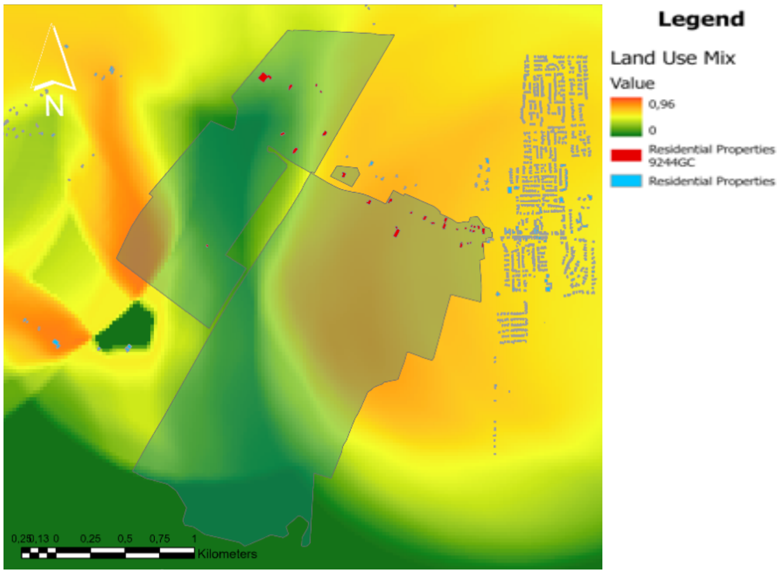


Figure 4.15. PC6 area 9244GC 1650-metre buffer

Figure 4.14 shows PC6 area 9713AB, a relatively large rural PC6 area. In both the Centroid- and PC6-approaches, this area receives a score of around 50. In the RP-approach, however, this score is over 85. This is mainly because in the RP-approach only the residential properties in the PC6 area are considered. In the Centroid- and PC6-approach, the large area to the east of the residential properties is included in the calculation, which means that a piece of land where nobody lives influences the WI-score. In the RP-approach, this problem no longer occurs, resulting in a more accurate score. The same applies to the PC6 area 9244GC, here too this PC6 area receives a higher value than in the other calculations because the large area to the south of the residential properties is no longer included in the calculation of the WI-score.

Furthermore, it could also occur that an area receives a lower WI-score in the RP-approach. Figure 4.16 shows that there are only four residential properties in this PC6 area, these properties are located in the part of the PC6 area with fewer components, the highest values can be found in the south. In the Centroid- and PC6-approach, the PC6 area receives a relatively high score based on the southern part. The RP-approach presents a more realistic scenario as the locations of the dwellings do not receive high scores. The same applies to figure 4.17, in this PC6 area which consists of two separate parts, only one residential property can be found, this dwelling is also located in an area with fewer high component scores than in the other parts of that PC6 area. Hence, this PC6 area obtained a higher score in other two approaches. However, the RP-approach is more accurate in this situation as it can be assumed that a WI is based on places where people live and not on zones where no one lives.

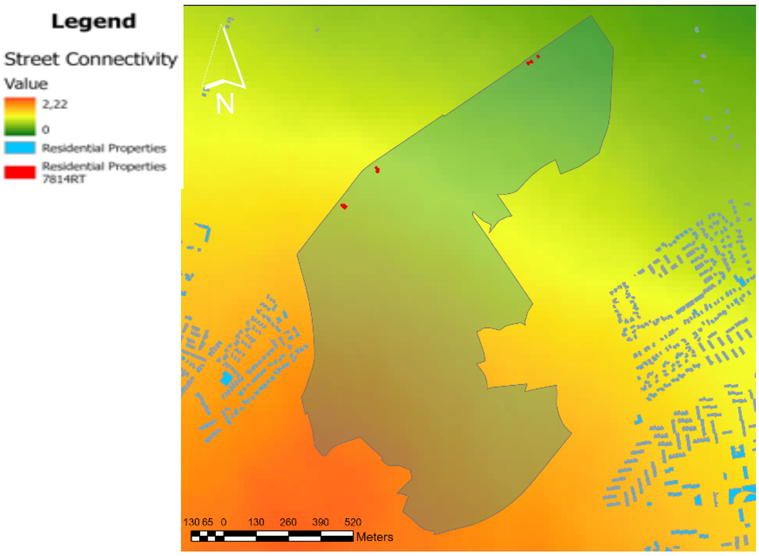


Figure 4.16. PC6 area 7814RT 1650-metre buffer

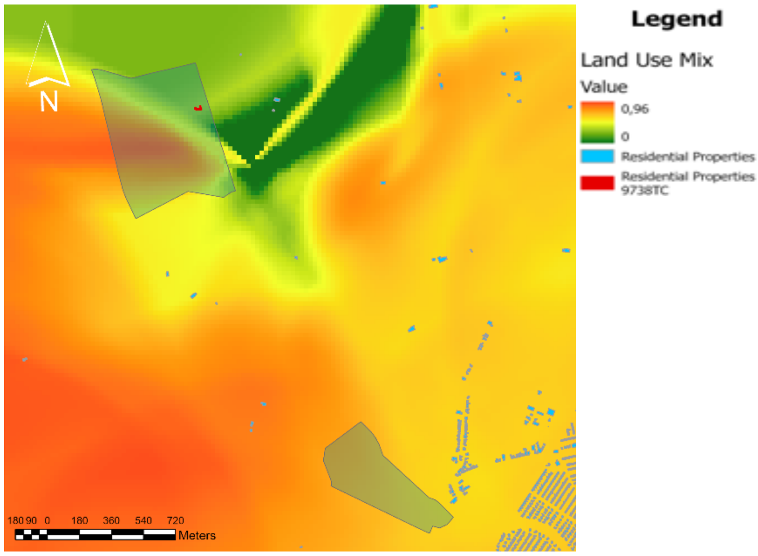


Figure 4.17. PC6 area 9738TC 1650-metre buffer

Conclusion

The 1650-metre approach shows different values than the 500-metre approach when comparing all three approaches. As the DoU decreases, the WI-score increases in both the PC6- and the RP-approach compared to the Centroid-approach. This is largely due to a buffer size that is more than twice as large as in the 500-metre approach. In addition, differences may arise as in these calculations there are about 7,000 PC6 areas less than in the 500-metre approaches. Despite these differences, the 1650-metre approach also shows that the RP-approach is no worse than the other two approaches. The visual examples show that, as in section 4.2, mainly the large rural PC6 areas receive a more accurate WI-score in the RP-approach. Thus, it appears that the differences between the two approaches are not very substantial, but the RP-approach generates the smallest margin of error.

4.4 Walkability, social behaviour and geography

The aim of this study is to combine health data with geographical data, or in other words to integrate health data at the PC6 level with the WI at the PC6 level. The size of a PC6 area can give a distorted impression of the WI-score. To reduce this level of inequality, two new approaches have been developed. In this chapter, these new approaches are compared with LifeLines health data and geographic data to understand the differences between the used approaches. Using regression analyses in SPSS and R-studio, the ways in which the three approaches in both 500m and 1650m affect socio-demographic and geographic factors are examined. A disadvantage of processing these results is that they cannot be exported directly from the LifeLines environment, as this is a protected environment. The results are therefore reproduced in tables. Because it is not necessary for every factor to compare all the approaches at all levels of urbanisation, the RP 500-metres approach is used primarily, as it has been shown to be no less reliable than the other approaches.

4.4.1 Degree of Urbanisation

The degree of urbanisation is divided into five categories from "very urban" (1) to "non-urban" (5). In the regression analysis conducted between urbanity and the WI, the dependent variable is the WI, and the degree of urbanity is the predictor or constant. This results in an increase by one class in the DoU (i.e. to less urban) to a decrease of the WI-score in all WI-approaches as can be seen in table 4.13.

In addition, there is a strong correlation between the DoU and the WI in all approaches (Centroid 1650 R =0.799, p =0.000; Centroid 500 R =0.799, p =0.000; PC6 1650 R =0.903, p =0.000; PC6 500 R =0.825, p =0.000; RP 1650 R =0.879, p =0.000; RP 500 R =0.816, p =0.000).

A remarkable aspect of these results is the difference between the Centroid-approach and the two new approaches. In the Centroid-approach, there is hardly any difference between the 500-metres and the 1650-metres approach, both drop by about 10 when the DoU increases by one. The PC6-approach and the RP-approach show better similarities; both approaches drop in the 500-metre approach approximately with a score of nine. However, in the 1650-metres approach both approaches show a strong increase by 12.708 and 14.196 respectively. This difference is most likely due to the chosen buffer size and does not indicate that one of the two distances predicts the WI better. Comparing the WI with the degree of urbanisation illustrates that both new approaches show a similar result, which indicates that based on all levels of urbanisation, there is no significant difference between the PC6- and RP-approaches.

Table 4.13. Results of linear regression modelling WI and DoU

Dependent variable	Predictor variable	Coefficient (B)	Constant	Adjusted R^2	P-value
Cent_500	DoU	-10.417	67.473	0.639	*0.000*
Cent_1650	DoU	-10.266	66.501	0.639	*0.000*
PC6_500	DoU	-9.666	60.174	0.680	*0.000*
PC6_1650	DoU	-14.196	87.382	0.816	*0.000*
RP_500	DoU	-9.266	58.377	0.667	*0.000*
RP_1650	DoU	-12.708	68.252	0.773	*0.000*

*Regression equation: Cent_500 = 67.473 - 10.417 * DoU. Cent_1650 = 66.501 – 10.266 * DoU. PC6_500 = 60.174 – 9.666 * DoU. PC6_1650 = 87.382 – 14.196 * DoU. RP_500 = 58.377 – 9.666 * DoU. RP_1650 = 68.252 – 12.708 * DoU.*

4.4.2 Walking time

Based on the previous results, it appears that the PC6-approach and the RP-approach do not differ significantly from each other based on all PC6 areas. However, it seems that the RP-approach is mainly effective in large rural PC6 areas, which is why a distinction was made between a degree of urbanisation of one and five during the analyses. The following analyses compare results from the LifeLines data with the WI at different levels of urbanisation. One of the questions in the LifeLines dataset concerns the number of minutes of walking per week in a participant's free time. The number of minutes of walking is compared to the WI-scores which measure what an increase in the WI does to the number of minutes of walking in leisure time. These analyses are conducted for all three approaches based on 500 metres. In table 4.14, the results of a regression analysis between walking time as dependent variable and the WI as predictor for a DoU of one and five can be seen.

Examination of walking for leisure and the degree of urbanisation in the Centroid-, PC6- and RP-approaches indicates that for an urbanisation of 1, i.e., strongly urbanised, the degree of urbanisation is not a significant predictor for the number of minutes walking for leisure (Centroid $R = 0.003$, $p = 0.837$; PC6 $R = 0.010$, $p = 0.420$; RP $R = 0.014$, $p = 0.282$). None of the approaches correlated strongly or were significant predictors of the number of minutes walked. At a DoU of 5 there is also no correlation, but all approaches are a significant predictor for the number of minutes of leisure walking (Centroid $R = 0.58$, $p < 0.001$; PC6 $R = 0.58$, $p < 0.001$; RP $R = 0.58$, $p < 0.001$). In the rural PC6 areas, the WI is a significant predictor for the number of minutes of walking for leisure, with the PC6- and RP-approaches showing the highest B; an increase on the WI of 10 results in approximately 17 minutes of additional walking for leisure. The fact that with a DoU of 5 the number of minutes of extra walking is more significant than with a DoU of 1 is partly since the range of WI scores in urban PC6 areas is larger than in rural areas.

Table 4.14. Results of linear regression modelling walking time and DoU

Dependent variable	Predictor variable	Coefficient (B)	Constant	Adjusted R^2	P-value
Wt Cent_500 (1)	DoU 1	-0.058	155.082	0.000	0.837
Wt PC6_500 (1)	DoU 1	-0.218	163.078	0.000	0.420
Wt RP_500 (1)	DoU 1	-0.285	166.241	0.000	0.282
Wt Cent_500 (2)	DoU 5	1.322	141.372	0.003	*<0.001*
Wi PC6_500 (2)	DoU 5	1.720	140.771	0.003	*<0.001*
Wi RP_500 (2)	DoU 5	1.763	139.767	0.003	*<0.001*

*Regression equation: Cent_500 (1) = 155.082 – 0.058 * DoU 1. PC6_500 (1) = 163.078 – 0.218 * DoU 1. RP_500 (1) = 166.241 –0.285 * DoU 1. Cent_500 (2) = 141.372 + 1.322 * DoU 5. PC6_500 (2) = 140.771 + 1.720 * DoU 5. RP_500 (2) = 139.767 + 1.763 * DoU 5.*

4.4.3 Age

The results of section 4.1 show that the older an individual becomes, the more he or she walks for leisure. The results also showed that the group up to 12 years old walks the most of all age categories, partly due to the lack of other means of transport. However, this group is not included in the analysis of the LifeLines dataset. That is as the minimum age to participate in their research is 18. In this results section, the Centroid 500-metres approach and the RP 500-metres approach are compared with the age of participants to determine whether the WI-score of a particular PC6 area can be a predictor of the age of its inhabitants.

As can be seen in table 4.15, there is no correlation between the WI and a participant's age in the Centroid-approach, but it can be seen that for all levels of urbanisation and for a DoU of 1, the WI is a significant predictor for age, except for a DoU of 5 (DoU all R =0.045, p <0.001; DoU 1 R =0.031, p =0.013; DoU 5 R =0.005, p =0.396). For all levels of urbanisation, the RP-approach and age are moderately correlated, although the WI is a significant predictor of age. A DoU of 1 shows a similar result, a DoU of 5 shows hardly any correlation and the WI is not a significant predictor of age (DuO all R =0.047, p <0.001; DoU 1 R =0.041, =0.001; DoU 5 R =0.005, p =0.350). Both approaches show a similar result where age in rural areas is not significantly explained by the WI, in urban areas and across the population it can be seen that as the WI improves, age decreases. However, this decrease is minimal and from the analyses it can be concluded that the WI is not relevant for the age of a participant based on the LifeLines dataset.

Table 4.15. Results of linear regression modelling Age and WI

Dependent variable	Predictor variable	Coefficient (B)	Constant	Adjusted R^2	P-value
Age (1)	Cent_500 DoU all	-0.034	50.617	0.002	*<0.001*
Age (2)	Cent_500 DoU 1	-0.045	47.047	0.001	*0.013*
Age (3)	Cent_500 DoU 5	-0.005	49.921	0.000	0.396
Age (4)	RP_500 DoU all	-0.040	50.610	0.002	*<0.001*
Age (5)	RP_500 DoU 1	-0.056	47.349	0.002	*0.001*
Age (6)	RP_500 DoU 5	0.007	49.759	0.000	0.350

*Regression equation: Age (1) = 50.617 – 0.034 * Cent_500 DoU all. Age (2) = 47.047 – 0.045 * Cent_500 DoU 1. age (3) = 49.921 – 0.005 * Cent_500 DoU 5. Age (4) = 50.610 – 0.040 * RP_500 DoU all. Age (5) = 47.349 – 0.056 * RP_500 DoU 1. Age (6) = 49.759 + 0.007 * RP_500 DoU 5.*

4.4.4 Gender

Section 4.1 has shown that women walk more than men; 60% versus 40%. In this section, it is investigated to what extent gender can be a predictor for the number of minutes walked, based on different levels of urbanisation. At each level, almost no correlation can be found, but gender is a significant predictor for the number of minutes walked (table 4.16). Across all levels of urbanisation, women walk about 20 minutes more than men during the week. At a DoU of 1 this is about 13 minutes but at a DoU of 5 this difference is more than 28 minutes (DoU all R =0.041, p <0.001; DoU 1 R =0.028, p =0.030; DoU 5 R =0.058, p <0.001). These results are consistent with the literature, notable is the large difference in the number of minutes walking between urban and rural communities. In this dataset 45,730 participants (58.9%) are female and 31,959 (41.1%) male.

Table 4.16. Results of linear regression modelling walking for leisure and gender

Dependent variable	Predictor variable	Coefficient (B)	Constant	Adjusted R^2	P-value
Walking for Leisure (1)	Gender DoU all	-20.758	196.136	0.002	*<0.001*
Walking for Leisure (2)	Gender DoU 1	-13.347	170.480	0.001	*0.030*
Walking for Leisure (3)	Gender DoU 5	-28.470	200.674	0.003	*<0.001*

*Regression equation: Walking for Leisure (1) = 196.136 – 20.758 * Gender DoU all. Walking for Leisure (2) = 170.480 – 13.347 * Gender DoU 1. Walking for Leisure (3) = 200.674 – 28.470 * Gender DoU 5.*

4.4.5 Educational level

As mentioned in sub question one, low educated people tend to walk more often than high educated people, mainly due to shorter commuting distances. But this applies mostly to big cities whereas these differences are less visible in the more rural parts of the country. It is interesting to investigate to what extent the WI plays a role in this phenomenon. In the LifeLines survey, a classification was made in nine levels of education from 'no education' to 'university education'.

By comparing the highest level of education to the WI, it seems that the WI is a significant predictor for the degree of education. This is consistent in both the PC6- and the RP-approaches, the Centroid-approach shows more outliers and is therefore less interesting (RP500 DoU all R =0.093, p <0.001; DoU 1 =0.082, p <0.001; DoU 5 R =0.015, p =0.006). In urban PC6 areas, individuals that live in a PC6 area with a high WI-score tend to have completed a higher level of education than individuals that live in PC6 areas with a lower WI-score. As the DoU decreases to rural PC6 areas, these differences become less present which indicates that in rural PC6 areas, the WI-score is less relevant for the highest obtained education of a participant. However, it should be noted that the number of participants varies greatly depending on the level of education, as can be seen in table 4.17. This could give a misleading result because individuals with a higher education tend to live more in urban areas. In table 4.18, the results of the linear regression between education and the RP500 WI are presented.

Education	Frequency
No education (did not finisch primary school)	454
Primary education (primary school, special needs primary school)	1,244
Lower or preparatory secondary vocational education (lts, vmbo)	9,792
Junior secondary education (mavo, vmbo-t)	10,955
Secondary vocational education or work-based learning pathway (mbo-long, bbl)	22,254
Senior general secondary education, pre-university (havo, vwo)	5,851
Higher vocational education (hbo, hts, heao)	20,079
University education	5,321
Other	1,213
Total	**77,689**

Table 4.17. frequency table highest attended education

Table 4.18. Results of linear regression modelling highest level of education and RP_500

Dependent variable	Predictor variable	Coefficient (B)	Constant	Adjusted R^2	P-value
Highest level of education (1)	RP_500 DoU all	0.011	5.160	0.009	*<0.001*
Highest level of education (2)	RP_500 DoU 1	0.012	5.696	0.007	*<0.001*
Highest level of education (3)	RP_500 DoU 5	-0.003	5.309	0.000	*0.006*

*Regression equation: Highest level of education (1) = 5.160 + 0.011 * RP_500 DoU all. Highest level of education (2) = 5.696 + 0.012 * RP_500 DoU 1. Highest level of education (3) = 5.309 – 0.003 * RP_500 DoU 5.*

4.4.6 Health

This research focuses on the relationship between walkability and health. This section examines to what extent the RP500-approach is a significant predictor of participants' health. Health in this case is measured by the Body Mass Index (BMI), the BMI is calculated by dividing the weight in kilograms by the square of the height in meters. (Voedingscentrum, sd). In the LifeLines dataset, this is categorised into three classes; normal (<=25), overweight (25-30) and obese (>30), in table 4.19, the frequencies are presented. As in previous sections, there is no correlation between the WI and BMI, however, for all levels of urbanity, the WI is a significant predictor for the BMI of participants (DoU all R =0.039, p <0.001; DoU 1 R =0.056, p <0.001; DoU 5 R = 0.020 p <0.001). In the analysis at all levels of urbanisation and at a DoU of 1, an increase in the WI-score leads to a decrease in BMI (-0.003), At a DoU of 5, an increase in the WI score leads to a slight increase in BMI (0.002). However, it should be noted that these values are so small that it is difficult to make statements about these outcomes as the differences are minimal, this is partly due to the fact that there are only three categories.

In addition to BMI, participants are asked to fill in their health status, whereby they can choose between 'excellent', 'very good', 'good', 'mediocre' and 'poor'. In this case, the extent to which health is a good predictor of the WI-score was examined. It is significant at all levels of urbanisation and at a DoU of 5 (DoU all R =0.009, p =0.013; DoU 1 R =0.022, p =0.088; DoU 5 R =0.017, p =0.002). The results are consistent with the expectation, namely that an increase in one degree of health, i.e. from 'good' to 'very good', corresponds to a higher WI-score. It can therefore be concluded that participants who consider themselves to have worse health also live less walkable neighborhoods. In table 4.20, the results of the linear regression between the RP500 WI and BMI and health are presented.

BMI Factor	Frequency
Normal	34,518
Overweight	30,536
Obese	11,837
Total	**77,689**

Table 4.19. Frequency table BMI

Table 4.20. Results of linear regression modelling RP_500, BMI and health

Dependent variable	Predictor variable	Coefficient (B)	Constant	Adjusted R^2	P-value
RP_500 (1)	BMI DoU all	-0.002	1.748	0.001	*<0.001*
RP_500 (2)	BMI DoU 1	-0.003	1.694	0.003	*<0.001*
RP_500 (3)	BMI DoU 5	0.002	1.701	0.000	*<0.001*
RP_500 (4)	Health DoU all	0.159	22.140	0.000	*0.013*
RP_500 (5)	Health DoU 1	-0.280	51.529	0.000	0.088
RP_500 (6)	Health DoU 5	0.164	11.084	0.000	*0.002*

*Regression equation: RP_500 (1) = 1.748 – 0.002 * BMI DoU all. RP_500 (2) = 1.694 – 0.003 * BMI DoU 1. RP_500 (3) = 1.701 + 0.002 * BMI DoU 5. RP_500 (4) = 22.140 + 0.159 * health DoU all. RP_500 (5) = 51.529 – 0.280 * health DoU 1. RP_500 (6) = 11.084 + 0.164 * health DoU 5.*

Conclusion

The analyses in this section have been conducted using the LifeLines dataset to determine if these results, combined with the WI approaches, correspond to the literature. In the analyses in which all three approaches are compared, it appears that the PC6- and RP-approaches provide more consistent results than the Centroid-approach and therefore correspond better to reality. In the analyses in which only the RP500-approach is used, mainly at a DoU of 5 the analyses show a significant result. It can be concluded that the RP-approach generates results consistent with the literature on walkability. For future analyses concerning walkability and health data, the RP-approach is most suitable when compared with the Centroid- and PC6-approach.

4.5 Analyses per WI component

The classification of the values of the raw components as provided by GECCO differ greatly. For the component green space the range falls between 0 - 100.029, for Land Use Mix between 0 - 0.9907, for Public Density between 0 - 258.982, for Public Transport between 0 - 4.33, for Retail Service Destinations between 0 - 84.43, for Street Connectivity between 0 - 3.43 and for Sidewalk Presence between 0 - 33.57.

To explain phenomena in a regression analysis with the components as independent variables, it is not a major problem that these mutual differences exist. However, for the interpretation, these different classifications make it more difficult to interpret the dependent variable in the same way. To compare the components more easily with each other, they are first standardised. This technique was already used in this study to create one WI from the components, but these standardised values are also used in the regression analyses. By standardising, the values of the independent variables (the components) no longer correspond to reality. However, by doing so, the components can be compared with each other and the effect on the dependent variable can be interpreted better.

4.5.1 Univariate regression

Table 4.21 shows the univariate regression analysis between the dependent variable; walking for leisure, and the standardised components. The variable 'walking for leisure' is taken from the LifeLines dataset, this variable shows the number of minutes walking for leisure (all walking movements except walking to and from work) per week. The WI 500-metres and each loose component is compared in a single regression with the number of minutes of walking for leisure, based on all PC6 areas with a DoU of 1 (urban) and PC6 areas with a DoU of 5 (rural). The full table can be seen in appendix II.

Dependent variable	**Predictor variable**	**Coefficient (B)**	**Constant**	**P-value**
Walking for Leisure (WfL)	RP_500	0.379	142.025	*<0.001*
WfL	Green Space	6.483		*<0.001*
WfL	Land Use Mix	8.238		*<0.001*
WfL	Public Density	-0.208		0.822
WfL	Public Transport	0.759		0.413
WfL	Street Connectivity	4.868		*<0.001*
WfL	Sidewalk Presence	4.443		*<0.001*
WfL	Retail Service Destinations	3.702		*<0.001*

Table 4.21. Results of univariate linear regression modelling Walking for leisure and single components

Based on Table 4.21, it can be seen that of the seven components, five have a significant influence on the number of minutes walking in leisure time. The components Public Density ($p = 0.822$) and Public Transport ($p = 0.413$) are not significant at a significance level of $p < 0.05$. A p-value lower than 0.05 can be considered statistically significant. This means that the components Public Density and Public Transport are not reliable enough to explain the number of minutes walking when measured over all levels of urbanity (Gelman & Stern, 2006). The other components are reliable enough to show that there is a significant relationship between them, and the number of minutes walked. The interpretation of the results works as follows: an increase of 1 on the WI 500 metres results in 0.379 more minutes of walking.

In other words, an increase of 10 on the WI results in 3.79 minutes more walking. In the same way, the other components can be interpreted. These univariate analyses were also conducted for urban and rural areas, resulting in clear differences; an increase of 1 in the component Sidewalk Presence results in 4.443 more minutes of walking for all PC6 areas. With a DoU of 1 (as can be seen in appendix II), the component is not significant and thus not a good predictor for the number of minutes walked. However, for a DoU of 5, the component is significant and an increase of 1 leads to 23 minutes more walking. This finding affects almost every component, at a DuO of 1 most components are not significant, at a DoU of 5 all components are significant and thus reliably predict the number of minutes walking.

4.5.2 Multivariate regression

In addition to univariate regressions, multiple regression analyses have been conducted in which all components are compared with each other to predict walking times. In order to use these results, it is first necessary to check for multicollinearity. This is relevant as no high correlation between the independent variables should occur; if so, the independent variables do not provide unique information but rather could be explained by each other, which leads to problems in interpreting the results. To test for multicollinearity, the Variance Inflation Factor (VIF) is used, which calculates the correlation and strength of the correlation between two predictor variables. It is generally considered that multicollinearity is present when the VIF is higher than five (Marcoulides & Raykov, 2019). In this study, none of the components has a VIF higher than 5, so it can be concluded that there is little or no multicollinearity. Table 4.22 shows the outcome of the multiple regression analysis in which the components predict the dependent value 'walking for leisure' in all levels of urbanity. This analysis was also conducted three times, once for all PC6 areas, once for the PC6 areas with a DoU of 1 and once for the PC6 areas with a DoU of 5, the full table can be seen in appendix II.

Dependent variable	Predictor variable	Coefficient (B)	Constant	P-value	VIF
Walking for leisure	Green Space, Land Use Mix, Public Density, Public Transport, Street Connectivity, Sidewalk Presence, Retail Service Destinations	3.545 5.679 -13.270 -0.674 2.986 11.653 -1.327	166.890	*0.001* *<0.001* *<0.001* 0.596 *0.021* *<0.001* 0.365	1.384 2.069 3.936 1.883 1.935 4.923 2.496

Table 4.22. Results of multiple linear regression modelling Walking for leisure and single components

Table 4.22 thus consists of a multiple regression analysis with seven components. The VIF differs per component but never exceeds five, although some components do come close. Only a few components are close to one and therefore do not or hardly correlate. It is notable that when both analysis are compared, Public Transport remains insignificant (p = 0.596) in the multiple regression while Public Density is significant at p = 0.021. However, in the multiple regression analysis, Retail Service Destinations no longer appears to be significant (p = 0.365). An additional noticeable aspect of the multiple regression is that the components are significant at all levels of urbanisation, however, when compared separately, only a few components remain significant, such as Green Space, Land Use and Public density.

Conclusion

Based on these tables, it appears that the degree of public transport is not an adequate indication for the duration of leisure-time walks. On the other hand, green space is very important in rural areas for the duration of leisure-time walks, as it increases significantly in both tables and is highly significant at $p < 0.05$. Both analyses show that the number of minutes walking increases for almost every component as the DoU decreases, this can be explained by several factors. First, a factor of influence is the fact that 6245 participants live in a PC6 area with a DoU of 1, compared to 33,765 participants in a PC6 area with a DoU of 5. Rural areas do not contain PC6 areas with a WI-score higher than 50, whereas urban areas contain PC6 areas with a score between 20 and 100. A score of 1 higher in an urban area is therefore less influential than in a rural area. Nevertheless, based on these analyses it can be concluded that the RP-approach and the separate components are reliable predictors of the number of minutes walking of the LifeLines participants.

5. Discussion

In this study, several choices have been made with respect to generating and comparing different methods for creating a WI. This chapter explains which other possibilities could contribute to better results. Hardware limitations and the absence of certain data, for example, resulted in the research not always progressing as planned. In this chapter, these missing variables are explained together with recommendations for policymakers and planners or a follow-up study.

Metropolitan bias
'Metropolitan bias', not to be confused with 'Urban bias', has been used to describe unequal allocation of recources by, for example, governmental services, the media, studies and polling to urban population areas over rural population areas (Ferre, Ferreira, & Lanjouw, 2010). This phenomenon is also evident in this study. Starting with the WI drawn up by GECCO, using a centroid to calculate the WI score. It appears that in urban areas it is less relevant if the Centroid-approach, the PC6-approach, or the RP-approach is used. Because of the different buffer sizes, the urban WI-scores do not differ that much from each other. The opposite appears to be true in rural areas, in large PC6 areas the Centroid-approach is far from accurate, yet GECCO chose to retain this method as it was considered accurate enough for calculating the areas in the Randstad. The fact that GECCO is based in Amsterdam could play a role in this regard.

Because the original WI was less suitable for rural areas, two new methods were presented in this study. These new methods were compared using the degree of urbanisation (DoU) based on CBS classification. The problem that occurred is that the distribution of the five classes is not equitable in the three Northern provinces. For example, in a DoU of five, there are almost 16,000 PC6 areas, which is almost half of all PC6 areas, while the PC6 areas with a DoU of one to three together consist of only just over 13,000 PC6 areas. This skewed ratio results in more unreliable results than if the different DoU consisted of approximately the same number of PC6 areas. This has not been implemented in this study, because the classification of CBS is used here.

In a follow-up study, it would be advisable to divide this classification more evenly and add new categories that further subdivide the urban PC6 areas. In the CBS dataset, PC6 areas with a surrounding address density of less than 500 addresses per km2 are the smallest possible scale, while a classification to less than 50 addresses per km2 could generate a clearer overview whereby not all rural areas fall into one category.

Missing LifeLines variables
Chapter 4.1 describes six demographic characteristics that mainly influence the use of active modes. These are age, gender, income, educational level, ethnic background and household composition. These characteristics were examined for the Netherlands and then compared with the LifeLines dataset in chapter 4.4. However, in the survey prepared by LifeLines, ethnic background and household composition are not properly reflected, there is no reference to ethnic background. The household composition is mentioned, but minimally, so that no good comparison can be made with the results from section 4.1. For this reason, these two characteristics have not been included in section 4.4. In a follow-up study it would be advisable to use a dataset in which these characteristics are also described properly.

In addition, it regularly happened that participants did not answer certain questions or that they misunderstood the question. For example, there were many outliers in questions about the number of weekly minutes walking and the distinction between walking for leisure and walking for work or transport was not always clear. For this reason, walking for transport has not been included and only calculations about the number of minutes walking for leisure have been made. Since most of the walking movements in the Netherlands are for leisure, this is not a major problem. However, in a follow-up study it is recommended to include all motives for walking in the calculations.

RUDIFUN dataset and household composition

In this study, the choice was made to use the BAG to draw up a better and more accurate method for calculating the WI. The BAG consists of all residential objects in The Netherlands and is therefore one of the most suitable methods to determine the WI-score based on residential locations. However, during this research, the PBL (Netherlands Environmental Assessment Agency) released a new dataset; the RUDIFUN dataset. With the RUDIFUN dataset (Spatial Densities and Function Mixing in the Netherlands), the connection between spatial densities, housing and function mixing on the one hand, and mobility, liveability and the real estate market on the other hand, can easily be examined. (Harbers, Amsterdam, & Spoon, 2022). Density and mixed use are central to this dataset and are underpinned by the FSI (Floor Space Index), GSI (Ground Space Index), MXI (Mix Use Index) and OSR (Open Space Ratio). These indicators for spatial densities connect seamlessly to the calculation of a WI.

Instead of calculating the RP-approach by means of the BAG, another approach could be used by means of the RUDIFUN dataset or in combination with the BAG. The RUDIFUN dataset is able to determine building types such as tower blocks, gallery blocks, medium-high-rise, terraced houses or bungalows. These different building types consist of different household compositions, something that was not considered during this research. In the RP-approach, every residential object counted as one, which means that a high-rise apartment has the same weighting as a single family house which could lead to misleading outcomes. The BAG dataset does show the number of people living in an object, but this number is not always correct and partly due to lack of time, this factor was not included in the calculation. In further research it is recommended to add a similar factor and with the RUDIFUN-dataset, different building types can be approached on the basis of numerical characteristics. In the GECCO calculation, population density is included based on the number of inhabitants per 100x100m grid cell, although using the RUDIFUN-dataset would be more accurate on a small scale.

Furthermore, land use mix has already been used by GECCO for its calculation of the WI, the factor was calculated by means of an entropy index which indicates the heterogeneity of land use. However, in their documentation document GECCO indicated that the entropy index has an major limitation. Although mixing or integration of land uses can change, entropy remains constant when distinct land-use types remain in constant relative proportions (Wagtendonk & Lakerveld, 2019). In their document Wagtendonk and Lakerveld already recommended to improve the land use mix component and avoid large buffer sizes as these effects are particularly visible in maps based on large buffer sixes. The RUDIFUN dataset could solve this level of inaccuracy as the MXI is calculated by dividing the Gross Floor Area for living by the total Gross Floor Area. This way, the ratio between living and non-living areas determines the MXI.

Weighing components

In this research, the WI is based on the components determined by GECCO. These components all gained the same weighting which means that, for example, population density and street connectivity count equally. However, in real life some components may be of more influence in the calculation of a WI than other components. Chapter 4.5 illustrated that all components predicted walking minutes with a different p-value and are therefore not homogenous. To provide a realistic WI it should be examined which components are more of influence and therefore should gain a different weighting. GECCO did not do this as this was time consuming, the same applies to this research but for further research it is advisable to examine different weightings for different target groups. For example, young people attach value to different factors than older people, older people may find the presence of green space more important than population density where younger people may find this the opposite. It is therefore interesting to generate a WI based on different target groups by using different weightings for different components. Providing a WI for different target groups makes the WI more useful in daily use because every individual values different factors related to walkability. Providing different weightings for different target groups will probably be difficult to examine but could lead to a much more realistic WI and will be very interesting for future walkability indexes.

Buffer or Service Areas
The calculation of the original WI was based on buffer zones around each grid cell location which indicates circular walk ranges, these buffer zones are based on straight-line Euclidean distances. However, a straight line is not optimal form a walkability standpoint because it doesn't take obstacles in account such as waterbodies, building blocks or other physical barriers. Wagtendonk and Lakerveld already mentioned in their technical document that a working with Service Areas would be a better approach for calculating a walkability score. Service Areas use the walkable infrastructure to calculate distances and are therefore more reliable than using a straight-line Euclidean distance. However, using Service Areas for a large area, such as three provinces, will take a lot of computing time as for every Service Area and for every component summary statistics have to be calculated. In further research, a WI should be calculated by means of Service Areas to ensure a more reliable index, although a limited number of residential locations is necessary due to limited computing power.

Comparing the approaches
In this study, different methods were used to calculate a WI. It appeared that in mainly rural areas, the Centroid-approach was not reliable enough to assign a WI-score per PC6 area. The PC6- and RP-approach seem to give a more reliable result in large PC6 areas. However, in this study it was not possible to statistically demonstrate if either the PC6-approach or the RP-approach is better for calculating a WI-score. The two approaches have been compared statistically, but mainly correlations have been found, which do not provide any information about the reliability of each approach. One of the aims of this research was to demonstrate which approach works best in rural areas, preferably by providing statistical evidence. However, partly due to lack of time and lack of statistical knowledge, this was not possible. However, the approaches have been compared visually. Based on the visualisations, it has been shown that the RP-approach in large PC6 areas is relatively more reliable than the PC6-approach. However, this was only investigated for the PC6 areas with the largest differences between the two approaches. In a follow-up study, it would be interesting to be able to statistically substantiate which of the two approaches best predicts the WI-score for all PC6 areas. This could be done by not relying on PC6 areas but by being able to work at small scales, these results would then have to be converted to PC6 areas but in this way a much more accurate calculation can be made. An interesting follow-up research would therefore be to combine the RUDIFUN dataset with Service Areas to calculate a WI, although this will be difficult to realise for large areas.

Hardware and software
As mentioned before, in the calculation of the 1650-metre approaches some data were lost, mainly due to hardware problems. In this research, ArcGIS Pro was used to calculate the new WI-approaches. Because a paid license is necessary to use all the functionalities of ArcGIS Pro, not all computers at the Radboud University are equipped with the program. The available computers are not that strong to compute analysis with large datasets, such as the BAG. The BAG consists of more than 15 million attributes which leads to large time-consuming analysis, sometimes up to more than 10 hours. In the PC6 approach waiting times also sometimes exceeded 5 hours but that could be done in one day. The RP-approach took most of the time, a lot more computing power and resulted in protracted analysis which couldn't be finished in one day. This led to a delay in the planning of this research. Because of these delays, the simplest solution was to cut up the datasets, perform the analysis separately and merge them again later. However, this led to some data losses and due to a lack of time there was no possibility to recalculate the 1650-metre analysis. Approximately 7,000 PC6 areas where lost which led to different outcomes than the 500-metre approaches. It is not known if these differences occurred

due to the data losses or by another phenomenon, but it is recommended to calculate all buffer sizes with the same amount of PC6 areas to get more reliable results in future research.

Recommendations for policymakers and planners

Most studies on walkability focus on urban areas, where there is a high degree of urbanisation and often several facilities are within walking distance. In rural areas this is considerably less true, which is why policy makers and planners in urban areas are more focused on walkability than those in rural areas.

However, this research has shown that most of the Netherlands consists of non-metropolitan regions, mainly rural areas with a low environmental address density per square kilometre. It has become clear that it is difficult to organise these often large rural areas in a way that is comparable to the city. This is often not possible since rural areas have fewer facilities and tend to be further apart than in urban areas. Nevertheless, policy makers and planners should give at least as much attention to rural areas as to urban areas when it comes to walkability. Chapter 4.1 has shown that in rural areas, the WI-score does not exceed 50 and that relatively more unhealthy people live in these areas than in urban areas. This has several causes, although one of them can be attributed to a sedentary lifestyle, which is greater in rural areas because, for example, people are more likely use motorized vehicles to get to a facility.

It is therefore important that policymakers and urban planners take a different view of the health of residents and how the built environment can contribute to that matter. One of the first recommendations is therefore to clearly identify the walkability situation in rural areas. The RP-approach could be a useful tool since it is based on the locations where people live, which is much more relevant in rural areas than in urban areas since the population density in rural areas is much lower than in urban areas. In addition, policymakers must consider the type of people that live somewhere. It has been shown that older, less educated, poorer and unhealthier people live in rural areas than in cities. These factors also reinforce each other, making this group increasingly large in rural areas.

This research has developed several methods by which walkability in rural areas can be better tested than with current approaches. Residents of rural communities are likely to value different factors than residents of urban communities. Therefore, a WI for rural areas is not the equivalent of a WI for urban areas. Policy makers and planners should make a proper assessment of which factors are relevant for rural areas, by giving weightings to certain components or by adding and removing components. In any case, a distinction should be made between WI in urban and rural areas because these two regions are significantly different from each other.

For each degree of urbanisation, it is necessary to examine which components are most relevant for the inhabitants of that area; after all, social status and other cultural factors have a major influence on people's behaviour. The built environment plays a major role in this respect, and it must therefore be designed in such a way as to highlight the most important components per degree of urbanisation. This is the only way to make a clear distinction between urban and rural areas and to build specifically for rural areas instead of applying urban techniques in rural communities. The methods in this study are a tool to determine the ultimate degree of walkability. However, a specific policy for rural communities with the most important components per target group or degree of urbanisation is necessary to overcome the metropolitan bias.

6. Conclusion

This chapter answers the three sub-questions and ultimately the main question. As the results section is fairly extensive, this chapter limits itself to the main findings from the results section.

What constitutes Walkability in social and cultural terms and between different degrees of urbanisation in the Dutch context?

The degree of walkability can be investigated in different ways. This study examined the socio-demographic characteristics that, according to the literature, mainly influence the use of active modes. This showed that there is no uniform Dutch walking culture but that the degree of active exercise varies per target group and per place. A large group responsible for walking are students, as they usually do not own motorised vehicles. Students mainly live in cities and in cities many amenities are within walking distance, therefore it is clear that in urban areas people walk more than in rural areas. People in the 65+ group also walk relatively much, mainly for recreational purposes. Income and education also play a role in the level of physical activity, which is partly caused by owning a car or not and the distance to work. Socio-demographic characteristics therefore have a certain influence on the use of active modes.

There is also a clear distinction between urban and rural communities. Rural communities are more often dependent on the car to reach similar amenities as residents in urban areas, and it has been shown that residents of rural areas are comparatively poorer and have more health problems than people in urban areas. These factors play a major role in the ability to walk every day and health problems are thus amplified as it drives a sedentary lifestyle. Walking distances also play a role in the transport choice of inhabitants, Dutch people do not like to walk long distances to find a public transport hub, for example. When distances are too long, the Dutch opt for other means of transport. Of course, the bicycle is very popular in the Netherlands and there is more of a cycling culture than a walking culture, especially compared to other countries.

The Dutch therefore mainly walk short distances (e.g., to the supermarket or into the city centre) or they walk for pleasure. Older people seek out regions such as South Limburg for recreational walking through the hilly countryside; for other purposes, they will more likely take a bike. The Netherlands is therefore not really a walking country and a clear walking culture does not exist, partly due to different target groups and the difference between urban and rural communities. It can be concluded that the Netherlands in terms of physical layout is very suitable for walking, but because of the many alternative transport options available (bicycle, public transport, car) there is no real walking culture in the Netherlands, rather a cycling culture.

What geographical unit is most suitable for conducting walkability studies to validate the index for all geographical areas in the Netherlands?

The purpose of this study is to demonstrate how the WI can best be integrated with health data. Because health data are privacy sensitive, the home address of a participant cannot be used as the responses could then lead back to the respondents. Therefore, only the postal code of a participant is known in the questionnaire. It appears that postal codes are not the best scale for visualising a walkability index, partly because there are large differences between different postal code areas. Rural postal code areas usually cover a larger area than postal code areas in urban areas, this disparity causes scores in rural areas to be less accurate than in urban areas.

However, because the WI must be linked to health data, a conversion must be made to a larger scale that is comparable across the country. Because health data and geographical data need to be linked in a GIS, a scale level is necessary that supports both datasets. Scale levels that are mainly used in GIS are provinces, municipalities, districts and neighbourhoods. These classifications were drawn up by the CBS and are mainly used in geographical analyses, as nearly all CBS and other public government data are available at these scales. In addition, postal codes can also be chosen as scale levels, which are subdivided into PC4, PC5 and PC6. PC4 means a postcode with four digits (e.g., 9244), PC5 means four digits + one letter (9244G) and PC6 means four digits and two letters (9244GC). PC6 is therefore the smallest scale level in postcode area, in urban areas these usually consist of only one or part of a street. The scale levels that can be used are therefore from large to small, country, province, municipalities, district, neighbourhood, postal code. To generate the WI as accurately as possible per location, it is advisable to use the smallest scale level, in this case PC6. During this research, no predetermined scale levels were found that could be imported into GIS. Of course, a researcher could draw up his own boundaries based on population density, for example, or by dividing the Netherlands into equal areas, 100x100 metres for example. However, this will be quite time-consuming, and it is questionable if health data at the PC6 level can be easily converted to self-created scale levels. In addition, the choice for self-created boundaries would have to be clearly substantiated, this time-consuming work does not occur with postal code levels.

The WI as drawn up by GECCO does not necessarily adhere to geographically delimited areas, as each grid cell is assigned a specific value, and this method of visualisation is therefore much clearer and more accurate than, for example, a large postal code area. During this research, the WI-score among dwellings was determined using the RP-approach; when these do not have to be converted to PC6 areas, this would also provide a convenient overview of the WI-score per dwelling. In this way, the exact score for each dwelling is known. When the WI does not need to be linked to health data, one of these two methods would be extremely suitable for compiling a WI.

However, when the WI does need to be linked to health data (this study), PC6 level is the most appropriate scale level. PC6 is easy to convert to other scale levels and visa versa, and people recognise postcode areas better than self-defined scale levels. Finally, when calculating personal data, postal code areas are the most suitable, because they cannot be traced back to the person and almost every inhabitant knows his postal code instead of his neighbourhood of district.

Which relations can be found between the walkability scores and individuals active mobility behaviour in different degrees of urbanisation and between different target groups?

It appears that the PC6- and RP-approaches both assign a better WI-score to PC6 areas than the Centroid-approach. When these approaches are compared statistically with socio-demographic characteristics, the results appear to be in line with the results from Chapter 4.1. According to chapter 4.1, age plays a major role in the number of weekly minutes walked, especially the young and the elderly walk relatively the most, unfortunately the survey of LifeLines has no data of persons younger than 18 years, so this comparison cannot be made very well. However, it was found that age does not play a significant role when compared to the WI. Almost all WI approaches show that as the WI-score increases, the age of a participant decreases, which could be explained by the fact that young people live more in urban areas than older people, this corresponds to reality. Also when looking at gender, this corresponds to reality, women walk more often than men.
Education is also a significant factor for the degree of walkability. The more educated an individual is, the more likely this individual lives in an area with a high WI-score. This is also in line with the results from chapter 4.1, people with a higher education live more often in urban areas and therefore live more often in an area with a better WI-score. The same is true for income, but no data were available for this field in the LifeLines dataset. However, the health status of participants is known in the LifeLines dataset, and this shows that people with a low or normal BMI live in areas with a better WI-score than people with a high BMI. The same applies to the degree of health of an individual; the less healthy a participant indicates they feel, the lower their WI-score will be. This is an outcome that was within the range of expectations but shows that urban planners and policy makers should pay more attention to

rural areas. People with poor health are more likely to live in areas with low WI-scores. The question is whether unhealthy people move here because it might be cheaper or whether they become unhealthy because they live in an area where a sedentary lifestyle is more common. This would be an interesting question for a follow-up study.

Finally, the walkability approaches show that a WI-score depends on the degree of urbanisation. Rural areas generally receive a lower WI-score (between 0 and 50) than urban areas (between 20 and 100) and as the degree of urbanisation decreases, so does the WI-score, which is true for all approaches. It can therefore be concluded that people with lower education, income and poorer health are more likely to live in areas with a low WI-score and thus mainly in rural areas. People who live in urban areas are generally healthier and therefore more likely to live in areas with a high WI-score.

To what extent can we evaluate and remodel the existing walkability index from GECCO to provide policymakers and city planners a more valid and reliable index for all Dutch geographical areas?

This study did not attempt to improve GECCO's walkability index by, for example, adding other components or removing certain components; GECCO's WI was kept as the basis. However, some adjustments were made to better integrate the WI with health data; this was done by GECCO based on a centroid, but it appeared that this approach was not reliable enough for rural areas.

In this study, two new approaches have been created, these have been made using the raw grid files of the components chosen by GECCO. The PC6-approach is the simplest approach and calculates the WI score for a PC6 area based on the total area instead of a centroid. This calculation is relatively easy to perform and does not take an extreme amount of computing power. This approach is more accurate and reliable for rural areas than the Centroid-approach, but it also has limitations. Certainly in large PC6 areas with relatively few inhabitants, the WI-score can still give a distorted impression due to high or low values of other components somewhere in the PC6 area.

The RP-approach does not have this problem; it is based on locations where people live and is therefore less sensitive to components further away in the PC6 area. This method does cost more time and computing power because the BAG is required to be used, but it certainly produces a more valid result for rural areas. The RP-approach is therefore the best way to adapt the existing WI so that it can be integrated with other datasets such as LifeLines. It can be demonstrated that rural areas often suffer from low WI-scores and that policy makers and urban planners should pay more attention to countering a sedentary lifestyle in rural areas. The PC6- and RP-approaches are two examples of ways to adapt the current WI and make it suitable for integration with other datasets. Of course, there are other possible ways and for follow-up research it would be interesting to explore the possibility of creating a new WI using the RUDIFUN dataset and/or Service Areas.

7. References

Bajracharya, B., Too, L., & Khanjanasthiti, I. (2014). Supporting active and healthy living in master-planned communities: a case study. *Australian Planner*, 359-361.

Cavill, N., Kahlmeier, S., & Racioppi, F. (2006). *Physical Activity and Health in Europe: Evidence for Action.* WHO Regional Office for Europe.

CBS. (2010-2014). *Onderzoek* . Den Haag: Centraal Bureau voor de Statistiek.

CDC. (2019). *Preventing Chronic Diseases and Promoting Health in Rural Communities.* Atalanta: National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP).

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

Chapman, D. (1996). *Creating Neigbourhoods and Places in the Built Environment.* Londen: Taylor & Francis.

Cristina, R., Bouldin, E., & Battista, R. (2021). Active living environments mediate rural and non-rural differences in physical activity, active transportation and screen time among adolescents. *Preventive Medicine Reports*.

CROW. (2021). *Inzicht in acceptabele* . CROW.

Cubukcu, E. (2013). Walking for sustainable living. *Social and Behavioral Sciences*, 33-42.

Daniels, R., & Mulley, C. (2011). Explaining walking distance to public transport: The dominance of public transport supply. *Explaining walking distance to public transport: The dominance of public transport supply*, 1-22.

de Haas, M., & Hamersma, M. (2019). *Loopfeiten.* Den Haag: Kennisinstituut voor Mobiliteitsbeleid (KiM).

Delis, C. (2017, 08). Flexibel openbaar vervoer, de oplossing voor ruraal Nederland? Faculteit der Managementwetenscappen.

Engelen, L., Dhillon, H., Chau, J. Y., Hespe, D., & Bauman, A. E. (2016). Do active design buildings change health behaviour and workplace perceptions? *OCCUPATIONAL MEDICINE-OXFORD*, 408-411.

Faskunger, J. (2013). Promoting Active Living in Healthy Cities of Europe. *JOURNAL OF URBAN HEALTH-BULLETIN OF THE NEW YORK ACADEMY OF MEDICINE*, 142-153.

Ferre, C., Ferreira, F., & Lanjouw, P. (2010). Is there a Metropolitan Bias? The Inverse Relationship Between Poverty and City Size in Selected Developing Countries. *The World Bank Economic Review, 26*(3), 1-34.

Fitzsimons, L. D. (2013, November). A multidisciplinary examination of walkability: Its concept, measurement and applicability. Dublin, Ireland: Dublin City University.

GECCO. (2022, 03 03). *www.gecco.nl*. Retrieved from www.gecco.nl

Gell, N., Rosenberg, D., Carlson, J., Kerr, J., & Belza, B. (2015). Built Environment Attributes Related to GPS Measured Active Trips in mid-Life and Older Adults with Mobility Disabilities. *Disability and Health Journal 8*, 290-295.

Gelman, A., & Stern, H. (2006). The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant. *the American Statistican, 60*(4), 328-331.

Grant, B. (2013). *Getting to great places.* San Jose: SPUR.

Hansen, A., Umstattd, M., Lenardson, J., & Hartley, D. (2015). Built Environments and Active Living in Rural and Remote Areas: a Review of the Literature. *Current obesity reports*, 484-493.

Harbers, A., Amsterdam, H. v., & Spoon, M. (2022). *RUDIFUN 2022: Ruimtelijke dichtheden en functiemenging in Nederland.* PBL Planbueau voor de Leefomgeving.

Haybatollahi, M., Czepkiewicz, M., Laatikainen, T., & Kytta, M. (2015). Neighbourhood preferences, active travel behaviour, and built environment: An exploratory study. *TRANSPORTATION RESEARCH PART F-TRAFFIC PSYCHOLOGY AND BEHAVIOUR*, 57-69.

Hiller, B., & Leaman, S. (1973). The man – environmentparadigm and its paradoxes. *Architectural Design*, 507-511.

IBM. (2020). *what is the formula for standardised (z) scores in SPSS*. Retrieved 7 1, 2022, from www.ibm.com/support/pages/what-formula-standardised-z-scores-spss

Johnson, B., Onwuegbuzie, A., & Turner, L. (2007). Towards a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 112-133.

Jun, H., & Hur, M. (2015). The relationship between waklability and neigborhood social environment: The importance of physical and perceived walkability. *Applied Geography*, 115-124.

Kukla, A. (2000). *Social constructivism and the phylosophy science.* London: Routledge.

Lang, J. (1987). *Creating Architectural Theory: The Role ofthe Behavioural Sciences in Environmental Design.* New York: Van Nostrand Reinhold.

Lau, K., & Tan, Z. (2018). Neigborhood Environment and Walking Behaviour in High-density Cities. *International Conference on Passive and Low Energy Architecture: Smart and Healthy Within the Two Degree Limit*, 839-842.

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*.

Liao, B., van den Berg, P., van Wesemael, P., & Arentze, T. (2020). Empirical analysis of walkability using data from the Netherlands. *Empirical analysis of walkability using data from the Netherlands*.

LifeLines. (2022, 03 06). *Over LifeLines*. Retrieved from Lifelins.nl: www.lifelines.nl

Marcoulides, K., & Raykov, T. (2019). Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods. *educ Psychol Meas, 79*(5), 874-882.

Mead, G. (1962). *Mind, Self and Society: From the standpoint of a social Behaviourist.* Chicago : University of Chicago Press.

Moore, D., Notz, W., & Flinger, M. (2013). The basic principles of statistics. In D. Moore, W. Notz, & M. Flinger, *The basic principles of statistics.* New York: Freeman.

O'Donoghue, G., Perchoux, C., & Mensah, K. (2016). A systematic review of correlates of sedentary behaviour in adults aged 18–65 years: a socio-ecological approach. *BMC Public Health*.

Owen, N., Cerin, E., Leslie, E., duToit, L., Coffee, N., Frank, D., . . . Sallis, J. (2007). Neighborhood walkability and the walking behavior of Australian adults. *American Journal of Preventive Medicine*, 387-395.

Pelfrene, E. (2010). *Standaardisatie van een statistische waarneming voor een of meerdere kenmerken.* Departement Kanselarij en Buitenlandse Zaken.

Raad voor de leefomgeving en infrastructuur. (2018). *De stad als gezonde habitat.* Rli.

Rafiemanzelat, R., Emadi, M., & Kamali, A. (2017). City sustainability: the influence of walkability on built environments. *Transportation Research Procedia*, 97-104.

Rao, V. (1980). Urban Bias and Rural Development. *Indian Economic Review, 15*(1), 75-83.

Robson, C. (2015). *Real World Research: Een bron voor sociale wetenschappers en praktijkonderzoekers.*

Roof, K., & Oleru, N. (2008). Public health: Seattle and King County's push for the built environment. *Journal of Environmental Health*.

Schaap, N., Harms, L., Kansen, M., & Wüst, H. (2015). *Fietsen en lopen: de smeerolie van onze mobiliteit.* Den Haag: Kennisinstituut voor Mobiliteitsbeleid.

Singh, R. (2016). Factors affecting waklability of neigborhoods. *Urban planning and architectural design for sustainable development*, 643-654.

Song, Y., Preston, J., & Brand, C. (2013). What explains active travel behaviour? Evidence from case studies in the UK. *ENVIRONMENT AND PLANNING A-ECONOMY AND SPACE*, 2980-2998.

Speck, J. (2012). *Walkable City.* Farrar, Straus and Giroux.

Stappers, N., Schipperijn, J., & Kremers, S. (2020). Combining Accelerometry and GPS to Assess Neighborhood-Based Physical Activity: Associations With Perceived Neighborhood Walkability. *Sage Journals*, 732–752.

Stilma, W., & Rijkenberg, S. (2021). *ONderzoek langs de meetlat.* Bohn Stafleu van Loghum.

Sugiyama, T., Leslie, E., Giles-Corti, B., & Owen, N. (2008). Associations of neighbourhood greenness with physical and mental health: Do walking, social coherence and local social interaction explain the relationships? *Journal of Epidemiology and Community Health* , 62-68.

Vale, D., & Pereira, M. (2016). Active accessibility: A review of operational measures of walking and cycling accessibility. *Journal of Transport and Land Use*, 209-235.

Van Thiel, S. (2007). *Research Methods in Public Administration and Public Management.* Routledge.

Vischer, J. C. (2008). Towards a user-centred theory of the built environment. *Building Research & Information*, 231-240.

Voedingscentrum. (n.d.). *BMI berekenen*. Retrieved 07 10, 2022, from www.voedingscentrum.nl/nl/bmi-meter

Wagtendonk, A., & Lakerveld, J. (2019). *Walkability score Netherlands version 1.0 Documentation of data and methods.* Amsterdam: Department of Epidemiology and Biostatistics.

Wagtendonk, A., & Lakerveld, J. (2019). *Walkability score Netherlands version 1.0; Dataset and technical documentation of data and.* Amsterdam: Geoscience and Health Cohort Consortium (GECCO).

Wang, H., & Yuqi, Y. (2019). Neighbourhood walkability: A review and bibliometric analysis. *Cities*, 43-61.

Wu, J., Wang, B., Wang, R., Ta, N., & Chai, Y. (2021). Active travel and the built environment: A theoretical model and multidimensional evidence. *Transportation Research Part D-Transport and Environment*.

Zuniga-Teran, A., Orr, B., Gimblett, R., Chalfoun, N., Guertin, D., & Marsh, S. (2017). Neighborhood Design, Physical Activity, and Wellbeing: Applying the Walkability Model. *INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH*.

8. Appendix

8.1 Appendix I Technical GIS document

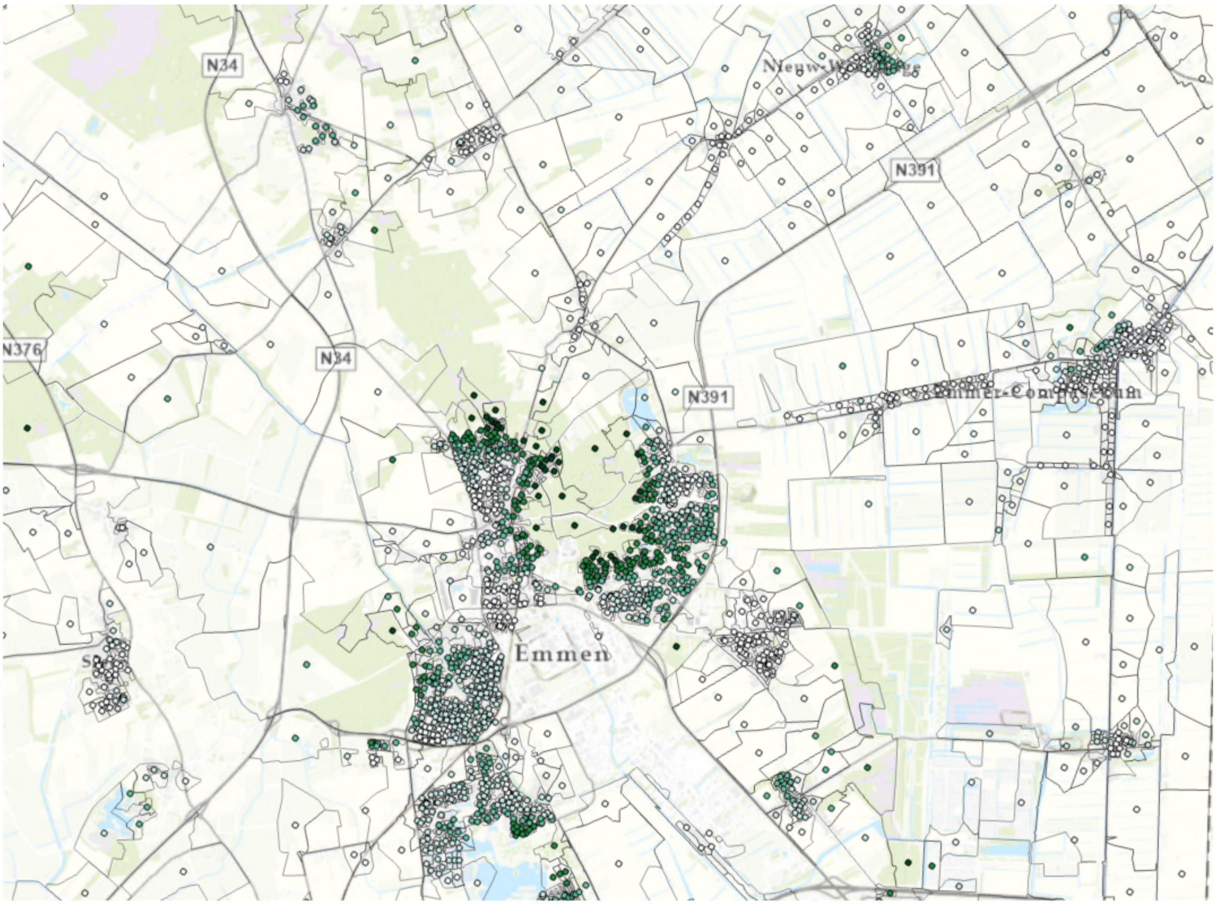
- For the calculations of the walkability index, the postcode 6 areas of 2018 were used. This dataset is publicly available and can be imported via ArcGIS or via the CBS website.
- These CBS PC6 areas have to be linked to the PC6 areas of LifeLines, as this file is not yet geo-referenced and therefore has to be linked to the CBS PC6 file by means of a 'join'. The result is dataset with all PC6 areas of LifeLines.
- Next, the raw raster files of GECCO are loaded into ArcGIS. These are the raster files of the seven components that make up the WI.

Three approaches have been used in this study; Centroid-approach, PC6-approach and the RP-approach. The calculations below must be conducted per component, both for 500 metres and 1650 metres. In total, this means 7 x 3 x 2 = 42 analyses (7x Centroid 500, 7x Centroid 1650, 7x PC6 500 etc.).

Centroid-approach

- The Centroid approach is calculated using the most central point of a PC6 area, this is an automatic function in ArcGIS.
- Extract to point: To convert grid values to points, points must first be generated. This is done by converting the polygons of the PC6 lifelines areas to points: PC6 - 'Feature to point'.
- Next, the tool 'extract values to point' is used. This tool extracts the cell values from a grid based on a point (Centre of postcode 6 area) and stores the values in the attributes table. Then, in the symbology tab, the cell values can be displayed by selecting 'RASTERVALU' at field.
- The raw raster data and the extracted points can now be overlaid, it will appear that the values of the points correspond to the values of the raster file. The result is a table per component with per PC6 area the Centroid value.
- This calculation is done 7 times (seven components) for the 500-metre analyses. When all calculations have been made, the tables can be exported to Excel and joined together in a table. Because not all components will probably have the same output, the function 'vertical look up' must be used in Excel. This function ensures that each PC6 area has the correct values of each component so that each unique PC6 area has the values of all seven components in the table.
- When each PC6 area has all the values of all the components, these values must be standardised. The standardised values go into the formula: 100 * ((value component - Min. value component) / (Max. value component - Min. value Component)). The result is the WI per PC6 area.
- The same steps should be carried out for the 1650-metre calculations.

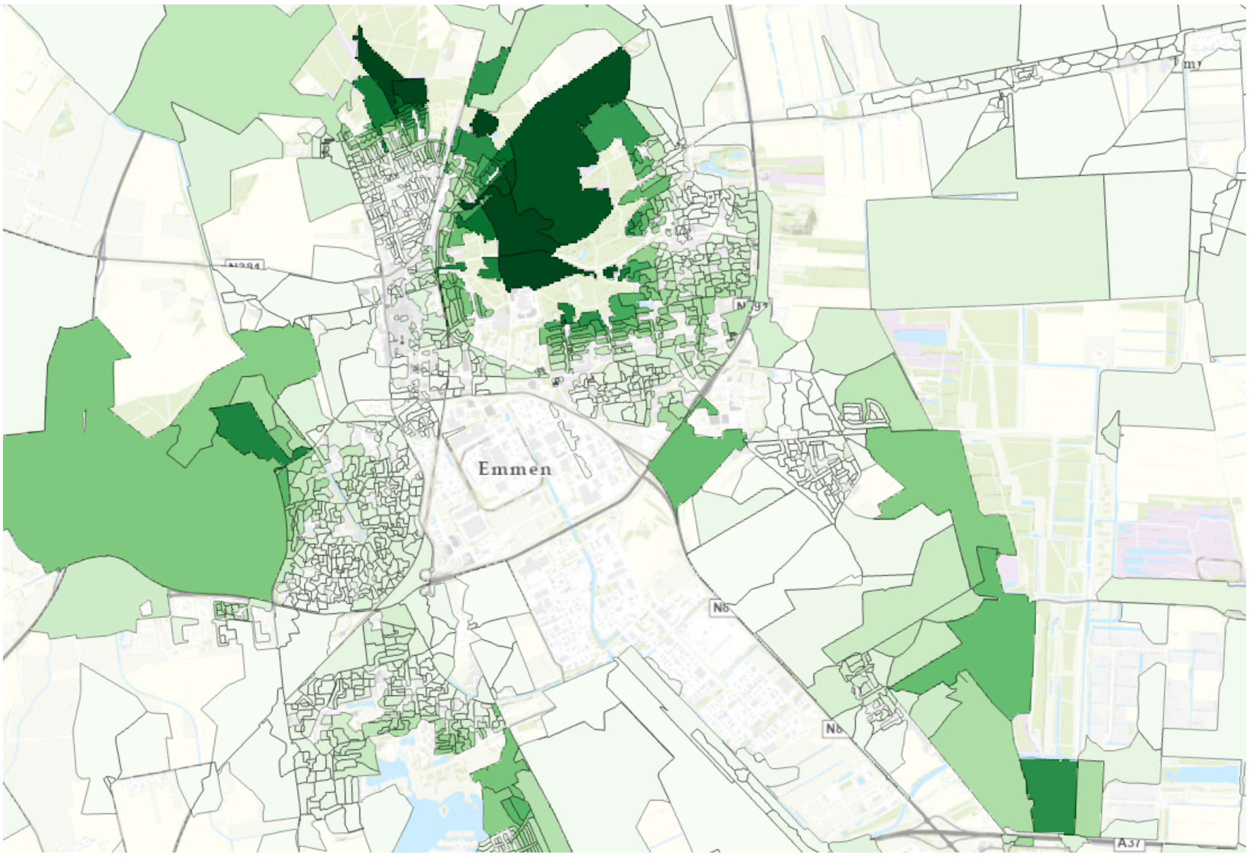
Central point taken from PC6 area and this gives the grid value under the point. Less accurate because this only says something about the location of the point and not about all the other areas in that same PC6 area.



PC6-approach

- - The PC6-approach is calculated based on all grid cells in a PC6 area, then averaged to determine the value for that PC6 area.
- - Zonal statistics as table: This calculation can be done by using the Lifelines PC6 shapefile in the 'Zonal Statistics as table' function. In this tool, the mean grid value under a particular postcode area is calculated and displayed in a table. For this function, the input feature is the Lifelines PC6 areas. The field that must be selected determines which object is taken, here the unique identification number (ID) must be selected so that each PC6 area has a unique name. The input value grid is the grid of one of the seven components, so this must be done seven times. Finally, you can choose what to show, in this calculation we chose mean because we want to know the average of all raster values in a PC6 area.
- This calculation is done 7 times (seven components) for the 500-metre analyses. When all calculations have been made, the tables can be exported to Excel and joined together in a table. Because not all components will probably have the same output, the function 'vertical look up' must be used in Excel. This function ensures that each PC6 area has the correct values of each component so that each unique PC6 area has the values of all seven components in the table.
- When each PC6 area has all the values of all the components, these values must be standardised. The standardised values go into the formula: 100 * ((value component - Min. value component) / (Max. value component - Min. value Component)). The result is the WI per PC6 area.
- The same steps should be carried out for the 1650-metre calculations.

Green space average grid value in PC6 area. The image below shows the average grid value in a PC6 area. All grid values within a PC6 area are summed and averaged, this average determines the value of the component in a PC6 area.

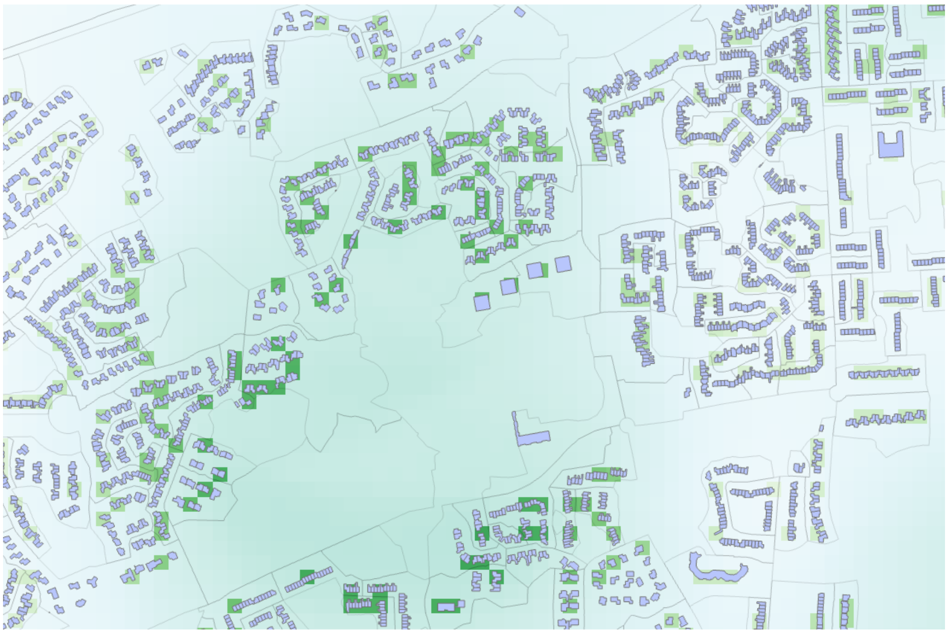


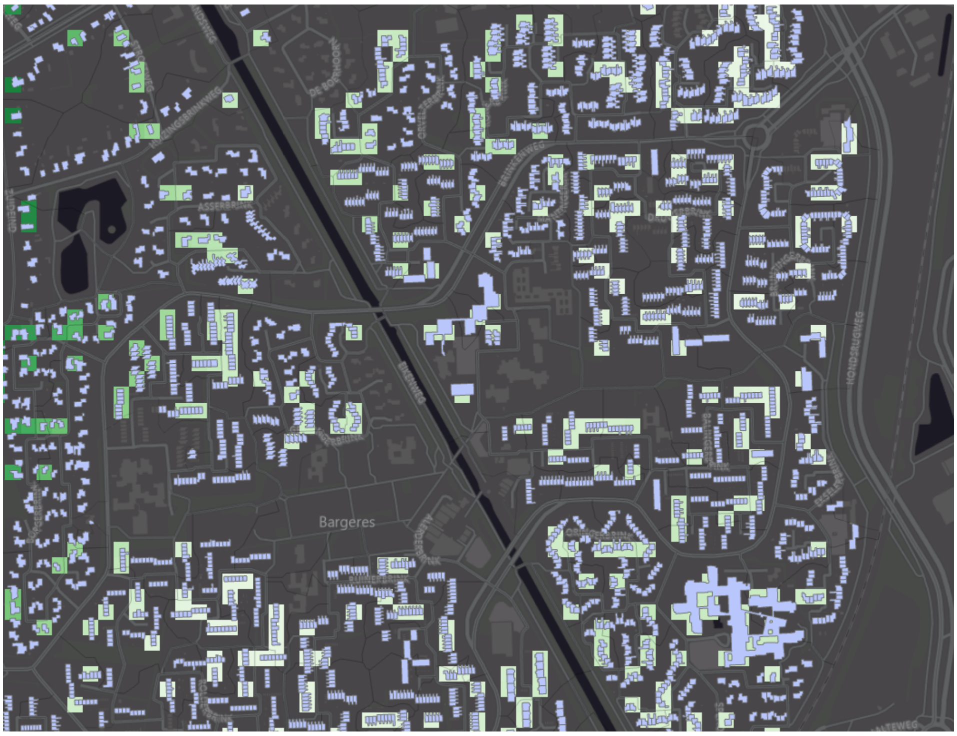
RP-approach

- - The RP approach is similar to the PC6 approach but instead of using all values in the PC6 area only the values under residential objects are used in the calculations using the public dataset; Basisregistratie Adressen en Gebouwen (BAG).
- - Import BAG properties (shape) into ArcGIS. Expression: Status = building in use and use residential function is equal to or greater than 1 (more than 1 residential objects in this building). Only the buildings with residential use remain, a new layer is created and this new layer is then 'clipped' with the Lifeline postcodes. Buildings with housing - clip - Lifeline PC6
- - Next, the same steps as for the PC6 approach are performed. The function 'Zonal Statistics as table' with as input the raster files and the BAG, the result is a table with per PC6 area the average of all raster values under a residential object.
- - Loading the BAG is very time-consuming and can only be seen on a small scale. It is therefore not easy to show an adaptation of the BAG on, for example, a provincial level, because it is too detailed, and this takes a lot of time and calculation work.
- This calculation is done 7 times (seven components) for the 500-metre analyses. When all calculations have been made, the tables can be exported to Excel and joined together in a table. Because not all components will probably have the same output, the function 'vertical look up' must be used in Excel. This function ensures that each PC6 area has the correct values of each component so that each unique PC6 area has the values of all seven components in the table.
- When each PC6 area has all the values of all the components, these values must be standardised. The standardised values go into the formula: 100 * ((value component - Min. value component) / (Max. value component - Min. value Component)). The result is the WI per PC6 area.
- The same steps should be carried out for the 1650-metre calculations.

BAG zonal statistics

The images below show all the residential properties with underneath (in green), the grid values per residential property. Next, the average of the grid values under the residential properties is calculated for each PC6 area and these determine the value for the PC6 area.





8.2 Appendix II regression walking for Leisure and components

Single regression regression for leisure and components (4.5)

Independent variable (Z-value)	Degree of urbanisation	Beta Score (*B*)	Significance (p)
WI_500_RP	*All*	*0.379*	*<0.001*
	1	-.285	0.282
	5	*1.763*	*<0.001*
Green Space	*All*	*6.483*	*<0.001*
	1	2.026	0.563
	5	*11.748*	*<0.001*
Land Use Mix	*All*	*8.238*	*<0.001*
	1	0.826	0.813
	5	*12.291*	*<0.001*
Public Density	All	-0.208	0.822
	1	-3.635	0.184
	5	*21.059*	*<0.001*
Public Transport	All	0.759	0.413
	1	*-4.840*	*0.010*
	5	*10.861*	*<0.001*
Street Connectivity	*All*	*4.868*	*<0.001*
	1	1.230	0.679
	5	*14.689*	*<0.001*
Sidewalk Presence	*All*	*4.443*	*<0.001*
	1	7.422	0.109
	5	*23.001*	*<0.001*
Retail Service Dest.	*All*	*3.702*	*<0.001*
	1	-0.528	0.785
	5	*24.545*	*<0.001*

Multiple regression for leisure and components (4.5)

Independent variable (Z-value)	DoU	Beta Score	Significance	VIF
Green Space	*All*	*3.454*	*0.001*	*1.384*
	1	2.143	0.666	2.020
	5	*7.285*	*<0.001*	*1.278*
Land Use Mix	*All*	*5.679*	*<0.001*	*2.069*
	1	0.926	0.888	3.515
	5	*6.313*	*<0.001*	*1.608*
Public Density	*All*	*-13.270*	*<0.001*	*3.936*
	1	-5.742	0.061	1.259
	5	-16.574	0.065	4.076
Public Transport	All	-0.674	0.596	1.883
	1	*-5.725*	*0.006*	*1.234*
	5	3.118	0.326	1.187
Street Connectivity	*All*	*2.986*	*0.021*	*1.935*
	1	-1.349	0.681	1.221
	5	6.744	0.053	2.783
Sidewalk Presence	*All*	*11.653*	*<0.001*	*4.923*
	1	*16.137*	*0.005*	*1.545*
	5	10.873	0.111	3.434
Retail Service Destinations	All	-1.327	0.365	2.496
	1	0.269	0.937	3.941
	5	6.811	0.169	1.409