GIMA
Geographical Information Management and Applications

The impact of Public Urban Green Infrastructure on health: a Amsterdam case-study

Jelle Verwer

Universiteit Utrecht
TU Delft
UNIVERSITY OF TWENTE.
WAGENINGEN UR

Master's Thesis 2020-2021
Geographical Information Management and Applications (GIMA)
Faculty of Geographical Sciences

Student information
Jelle Verwer
6557511
j.verwer@students.uu.nl

Universities
University of Utrecht
Wageningen University & Research

Under Supervision of
Dr.ir. R.J.A (Ron) van Lammeren (WUR)
Dr.ir. L (Lammert) Kooistra (WUR)

Submission date
02-03-2021

Summary

The Impact of Public Urban Green Infrastructure on Health: a Amsterdam Case Study

The objective of this study is to quantify the relation, or relationships, between public urban green infrastructure (PUGI) and citizens' health in the municipality of Amsterdam. The results show what implications these relationships have on the (planned) PUGI of the year 2050. It is argued in many studies that increased proximity of PUGI has a significant positive impact on, among other things, the prevalence of obesity, loneliness, and mental health problems. Amsterdam has a different urban-fabric than the cities that are the focal point of existing studies (mainly north-America). The null-hypothesis is that within the municipality of Amsterdam the proximity of PUGI is less a determinant for health than in existing studies. Among the reasons to assume this is the skew division of PUGI in Amsterdam, with high-income (central) areas in general having less PUGI than lower-income (outer) areas.

This study uses a multiple regression model to substantiate and quantify this relation or the null-hypothesis. Urban neighbourhoods, from which there are 91 in Amsterdam, are the aggregation-unit of choice; and beside PUGI several covariables are considered. The findings show that PUGI has no significant-correlation with obesity, mental health problems and loneliness. The covariables, economic-status and neighbourhood design are important explanatory variables for the health-variables. The analysis of the 2050-PUGI showed that the interventions the municipality propose strongly aim to increase the inclusion of low-SES people into the PUGI. A push which is much needed knowing that low-SES citizens are less healthy in general and make less use of PUGI. The 2050-policy additionally tries to increase the quality of the PUGI by introducing park-concierges, sport-amenities, social events, and functional-diversification of the parks (depending on the neighbourhood it is in).

In conclusion, the PUGI that Amsterdam neighbourhoods have in their proximity is currently an insignificant determinator for health. Future studies could be helped by introducing new variables such as fast-food-outlets and cardiovascular diseases, which are currently not openly available on the neighbourhood-level. In addition, the 91 neighbourhoods may be too small of a sample-size for robust statistical modelling; leading to significant outliers within the sample-size that are a nuisance to the model-fit and coefficients. Using a (fine-grained) tessellated aggregation in combination with data on this level would allow for better statistical modelling and more robust outcomes. This data-precision is unavailable for Amsterdam, substantiating the choice for using the neighbourhood as aggregation.

Abbreviations

BMI	*Body Mass Index*
CBS	Central Bureau of Statistics
CSV	Comma Separated Value
FME	Feature Manipulation Engine
FSI	Floor Space Index
GDP	Gross Domestic Product
GI	Green Infrastructure (includes blue infrastructure)
GSI	Ground Space Index
GWR	Geographically Weighted Regression
LISA	Local Indicators of Spatial Association
MAUP	Modifiable Area Unit Problem
MIF	MapInfo Interchange Format
NGO	Non-Governmental Organisation
NUA	New Urban Agenda
OLS	Ordinary Least squares
$PM_{(10)}$	Particulate Matter (2.5-10 micrometre)
POS	Public Open Spaces
PUGI	Public Urban Green Infrastructure (includes green, blue, cycle & public transport)
RIVM	National Institute for public health and the environment
RQ	Research Question
SDG	Sustainable Development Goals
SES	Socio-Economic Status
SHP	Shapefile
TOP10NL	Topographic map (ideal for 1:5.000 - 1:25.000 visualisation)
UN	United Nations
WGS84	World Geodetic System (84')
WHO	World Health Organization

Contents

1. Introduction 7
2. Urbanisation, PUGI And Health in Amsterdam 10
2.1 Introduction 10
2.2 PUGI Trends 10
2.3 Spatial Extent 11
2.4 PUGI Policy 12
2.5 Conclusion 14
3. Theoretical Framework 15
3.1 Introduction 15
3.2 PUGI in Relation to Health 16
3.3 Covariables 19
3.3.1 Introduction 19
3.3.2 Socio-Economic Status 19
3.3.3 Urban Density 20
4. Methodology 22
4.1 Introduction 22
4.2 Literature Review 23
4.3 Spatial Scope 23
4.4 Data Collection 24
4.5 Data Preparation 25
4.6 Data Filtering 26
4.7 Data Aggregation 26
4.8 Multiple-regression analysis 26
4.8.1 Model Types and Indicators 26
4.8.2 Spatial Autocorrelation 27
4.8.3 Modifiable Area Unit Problem 28
4.8.4 Ecological Fallacy 29
4.8.5 Edge Effects 29

4.8.6 Validation 29
4.8.7 Spatial Patterns 30
4.8.8 Standardization 30
4.8.9 Matrices 30
4.9 The 2050 PUGI policy 32
4.9.1 Introduction 32
4.9.2 Comparing 2020 And 2050 32
4.9.3 Expected Health impact of 2050 PUGI 33
5. Analysis 34
5.1 Introduction 34
5.2 Research Question 1 34
5.2.1 Introduction 34
5.2.2 Overweight, Obesity and BMI 35
5.2.3 Mental Health Problems and Loneliness 36
5.2.4 Healthcare Spending 36
5.2.5 Covariables 37
5.2.6 Variable Overview 37
5.3 Research Question 2 38
5.3.1 Introduction 38
5.3.2 Data Collection 39
5.3.3 Data Aggregation 41
5.3.4 Data Standardization 42
5.3.5 Health Variables 42
5.3.6 Covariables 43
5.4 Research Question 3 46
5.4.1 Introduction 46
5.4.2 Variable Exploration 47
5.4.3 OLS-model: Obesity 48

5.4.4 OLS-model: Obesity (K5) 50

5.4.5 Spatial-Lag-Model: Obesity (Q1) 50

5.4.6 Buffered Models: Obesity 51

5.4.7 OLS-Model: Mental Health Problems (K5) 52

5.4.8 Buffered Models: Mental Health Problems 54

5.4.9 OLS-Model: Loneliness (K5) 55

5.4.10 Spatial Error-Model: Loneliness (Q1) 57

5.4.11 Buffered Models: loneliness 58

5.4.12 Multiple-Regression Summary 59

5.5 Research Question 4 60

5.5.1 Introduction 60

5.5.2 2050 Policy Implementations 61

5.5.3 2050 PUGI Strongpoints 62

5.5.4 2050 PUGI Weak points 65

5.5.5 Conclusion 66

6. Conclusion 68

7. Recommendations and discussion 70

8. Bibliography 71

1. Introduction

Urban expansion and urban densification are evidently related to changes within the social, economic and *health* domains of urban citizens. The *health* domain, which includes among other things obesity, cardiovascular diseases, mental health problems, loneliness, and depression, is a growing concern. The urbanization process, that causes these increasingly larger and more dense urban environments, takes place due to the many upsides newcomers associate with cities (Murali et al., 2018). But the influx of newcomers and thereof coming urban densification, urban expansion, and green infrastructure intensification, have a negative impact on the availability and proximity of green infrastructure. Growing urban areas in general have a tendency towards increasingly less *public* urban green infrastructure (PUGI), in absolute as well as relative numbers. The terminology of *PUGI is* based upon the concept of *urban green infrastructure (UGI)* from (Norton et al., 2015). In addition, urban environments on average currently already have (slightly) more physical and mental health problems compared to more rural environments (Peen, Schoevers, Beekman, & Dekker, 2010). For the Netherlands, it is demonstrated that unhealthy lifestyles and thereof coming physical and mental problems manifest themselves more in larger cities compared to smaller cities and rural areas, as a recent study from Arcadis shows (Boon, Westerink, Noten, van der Schee, & Derksen, 2020).

The availability of nearby and publicly available green infrastructure, *public urban green infrastructure* (PUGI), may be one of the solutions to mitigating and resolving these mental and physical urban health problems. Bosch & Goossen give a twofold positive effect PUGI has on urban citizens: *'Well planned and designed green infrastructure, including water and soil, can contribute to climate change adaptation and at the same time promote and support healthy urban living'* (Bosch & Goossen, 2016, p. 12). Due to global warming, the negative health effects coming from PUGI-scarce urban environments, is increasing. At least partly due to more extreme weather phenomena with serious health hazards that can be (partly) mitigated by the presence of PUGI (Norton et al., 2015). For example, PUGI improves flooding and heat mitigation by the use of green roofs, parks, and urban water bodies, and in addition PUGI creates shadow-rich and lush environments that lead to more comfortable and healthier urban-living conditions during these extreme weather phenomena.

There are many scientific studies into the *PUGI* ↔ *HEALTH* relation in miscellaneous case study-areas. In for example Oslo (Mouratidis, 2020), Melbourne (Gunn et al., 2017) and Barcelona (Mueller et al., 2020) there is a positive relation between increased PUGI-presence and a decrease in the prevalence of (some) health problems. These studies however are not transferrable to the Amsterdam-area due to the importance of local factors in this relation, something that is mentioned in for example studies from Li and Staatsen (T. Li et al., 2020; Staatsen et al., 2017). In Amsterdam, every 5 years PUGI usage is monitored and PUGI-users are asked how they perceive the green-infrastructure, last time in 2018 (OIS, 2018). The 2018 outcomes show a changing usage-pattern compared to the 2013-monitor. For example, parks are less used for physical activity compared to 2013 due to sporters being put off by the large crowds of people. Park usage

increased from 89% in 2013 to 93% in 2018, neighbourhood green usage rose even quicker from respectively 54% to 61% of Amsterdam citizens. A trend that is accompanied and amplified by an absolute population increase.

Because urban spatial patterns and their relations, like PUGI and their relationship with health variables are dependent on the specific context of an urban area (*e.g.* the past *economic, cultural, densification* and *expansion related* developments) (World Health Organisation, 2010), a case study, examining the relation between PUGI and health-variables in the municipality of Amsterdam is crucial in understanding the local patterns and relations. A case study on Amsterdam PUGI and health-variables has added value for the Amsterdam policy makers like the municipality (Dutch: *gemeenteraad)* and district councils (Dutch*: stadsdeelraden).* This emphasizes the societal significance for the municipality of Amsterdam, and above all the Amsterdam citizens. The scientifical relevance comes from the fact that many previous studies on this subject use different case study areas, and Amsterdam has unique features that were not considered in this existing body of research.

The PUGI and health variables that are used in this study have an expected significant (spatial) relation based upon existing literature, but can also be unique to Amsterdam, and based upon local factors that are expected to be of additional importance. To strengthen the relevance of the research, the scope is extended into the future; meaning an added research goal is reflecting on a future scenario for the urban green infrastructure in Amsterdam (± 2050), using the 2020 analysis outcomes. And by this means, try to analyse what impact this future infrastructure has on a selection of health variables. To improve the transparency, quality and reproduction of this research, the data used will solely consist of open datasets provided by renowned Dutch governmental institutions.

The central problem statement that grasps the aim of this study is: '*The case of Amsterdam neighbourhoods: what public urban green infrastructure variables impact health variables, and how can future green infrastructure interventions change these relationships?*' There are 4 research questions that help answer this problem statement in a structural and logical manner:

1. *What theoretical relationships are found between PUGI-variables and health-variables?*
   - Review of existing related research and literature to find what variables and past findings are of importance.
2. *What PUGI-variables, health-variables and co-variables are openly accessible (on a neighbourhood level in Amsterdam)?*
   - Obtain relevant open-data using a selection of geo-portals
   - Pre-process this data into usable multiple regression variables
3. *What statistical relationships are found between PUGI and health variables among neighbourhoods in Amsterdam?*

- Using a multiple regression analysis, what relation can be found between the independent PUGI variables, the dependent health variables, and the covariables.
- How can the findings be interpreted and validated, statistically and/or theoretically.

4. *How can future public green infrastructure impact these relationships between PUGI-variables and health-variables?*
   - Review, categorize and summarize the 2050 PUGI interventions
   - Analyse the 2050 PUGI interventions based upon the 2020-regression outcomes.
   - What interventions are expected to positively impact health the most and what are the shortcomings of the 2050 PUGI

The hypothesis is that within a multivariate spatial regression model, more PUGI does only slightly significantly improve health variables in Amsterdam. It is expected that important covariables, like the socio-economic status and urban density, have a larger and more significant impact on health variables, compared to PUGI. This hypothesis is in line with a research done in Utrecht where *'Positive relations between health and green spaces are sometimes corrected for other factors influencing health'* (Bosch & Goossen, 2016, p.90). A unique characteristic of Amsterdam (and Dutch cities in general), that increases scientific relevance for a Amsterdam-specific study, is that its citizens move around by bike, foot, and public transport relatively a lot compared to cities outside of the Netherlands, where the car is often the preferred choice of (inner-city transport), for this reason the location of bike paths and public-transport stops are additional variables of interest for this case-study (*table 1-1)* (Harms & Kansen, 2018).

Table 1-1: Share of transport-modes in Amsterdam for work-trips in 2016 (Harms & Kansen, 2018)

Mode of transport	*Bicycle*	*Car*	*Public transport*	*Train*	*Other*
Share (%)	48%	21%	16%	1%	14%

It is expected that, for the 2050 PUGI-scenario, small changes in neighbourhoods with currently little PUGI will lead to relatively more benefits than adding green to already green neighbourhoods, a trend also seen in (Paulin, Remme, & de Nijs, 2019). In addition, the quality of future green infrastructure may be more important than the quantity, as Amsterdam has much PUGI that is not extensively used right now, the incentive to go there is low compared to some smaller PUGI, due to missing attractive amenities, accessibility, and good maintenance.

The research will start with an overview of the current PUGI-developments in Amsterdam, the relevance of this city, its PUGI-policy, and unique characteristics (Chapter 2). Hereafter, theory surrounding PUGI in relation to different health variables will be elaborated (Chapter 3). Subsequently the analysis steps per research question (RQ) will be elaborated in the methodological chapter (Chapter 4), and hereof coming the analysis outcomes are elaborated per RQ (chapter 5). Finally, the concluding remarks, recommendations and reflection are communicated (chapter 6, 7 & 8).

2. Urbanisation, PUGI And Health in Amsterdam

2.1 Introduction

Within the Netherlands, Amsterdam is the 1st most populous and 4th most dense municipality (*after The Hague, Leiden, and Haarlem*). Simultaneously Amsterdam has the highest *absolute* urbanisation rate; between 2016 and 2025 Amsterdam will be expected to grow with an additional 70.000 inhabitants to reach approximately 900.000 inhabitants in 2025 (Paulin et al., 2019). In this same time-frame an additional 52.500 housing units are expected to be built within the municipal boundaries to accommodate these people (Gemeente Amsterdam, 2016). Increasing population density, housing density and re-zoning for more housing therefore is the medium to long-term policy of the municipality as written in the 2040 vision (Zanen, Ponteyn, & Keijzer, 2011). Due to rezoning for housing and building all 52.500 housing units within the municipal borders, the population increase results in more intensive usage of the existing PUGI within the municipality of Amsterdam, and sequentially less available green per inhabitant (Paulin et al., 2019). This can be problematic for urban dwellers who use the PUGI as an catalyst for their social life, physical activity, and mental wellbeing.

2.2 PUGI Trends

Due to this steady population growth, resulting mostly from a positive net migration, the public parks in Amsterdam, an important part of the PUGI, will see an expected annual linear increase in usage of between 3% (small neighbourhood parks) and 9% (large urban parks) (Paulin et al., 2019). This in return reduces the available PUGI per person, PUGI that is used for physical activity and social life. PUGI per person decreased slightly from 71.4 M^2 in 2015 to 70 M^2 in 2016 but hereafter it went down more quickly to 60 M^2 of PUGI per person in 2020. Twofold cause for this trend is the strong population increase in combination with an absolute PUGI *decrease* in the same period (De Gezonde Stad, 2021).

In addition to creating lush environments for Amsterdam citizens that provide space for physical activity, mental wellbeing and social interaction, the proximity of PUGI plays an small role in reducing air pollution from (mainly > PM_{10} particles) (Chen & Kan, 2008; Wesseling, Beijk, & Kuijeren, 2008). PUGI in addition increases on average the property values in its proximity (KPMG, 2012), in theory the increased property-taxes could finance the investment in PUGI. Regarding mental health, sufficient nearby public urban green is proven to reduce stress, loneliness and depression (Bos, Horjus, Visser, & de Wilde de Ligny, 2020; Fan, Das, & Chen, 2011; Holtan, Dieterlen, & Sullivan, 2015; Jennings & Bamkole, 2019). This combined expected effects the proximity of PUGI has on a dwellers' physical health, social life, productivity, and mental stability make this a societal relevant topic too look deeper into.

2.3 Spatial Extent

The municipal borders of Amsterdam are the spatial extent of this study, the 99 neighbourhoods are the aggregation units within this municipal extent. The spatial extent and the underlying aggregation units are important for data-collection, as they both introduce limitations to what data can be used. Especially the neighbourhood aggregation poses some limits to the usage of some variables that are not available on this scale-level. The current (2020) PUGI (+ bicycle infrastructure) is shown in *figure 2-1* to create a sense of the extent of the research- area and the spatial patterns of the green, blue and cycle infrastructure.

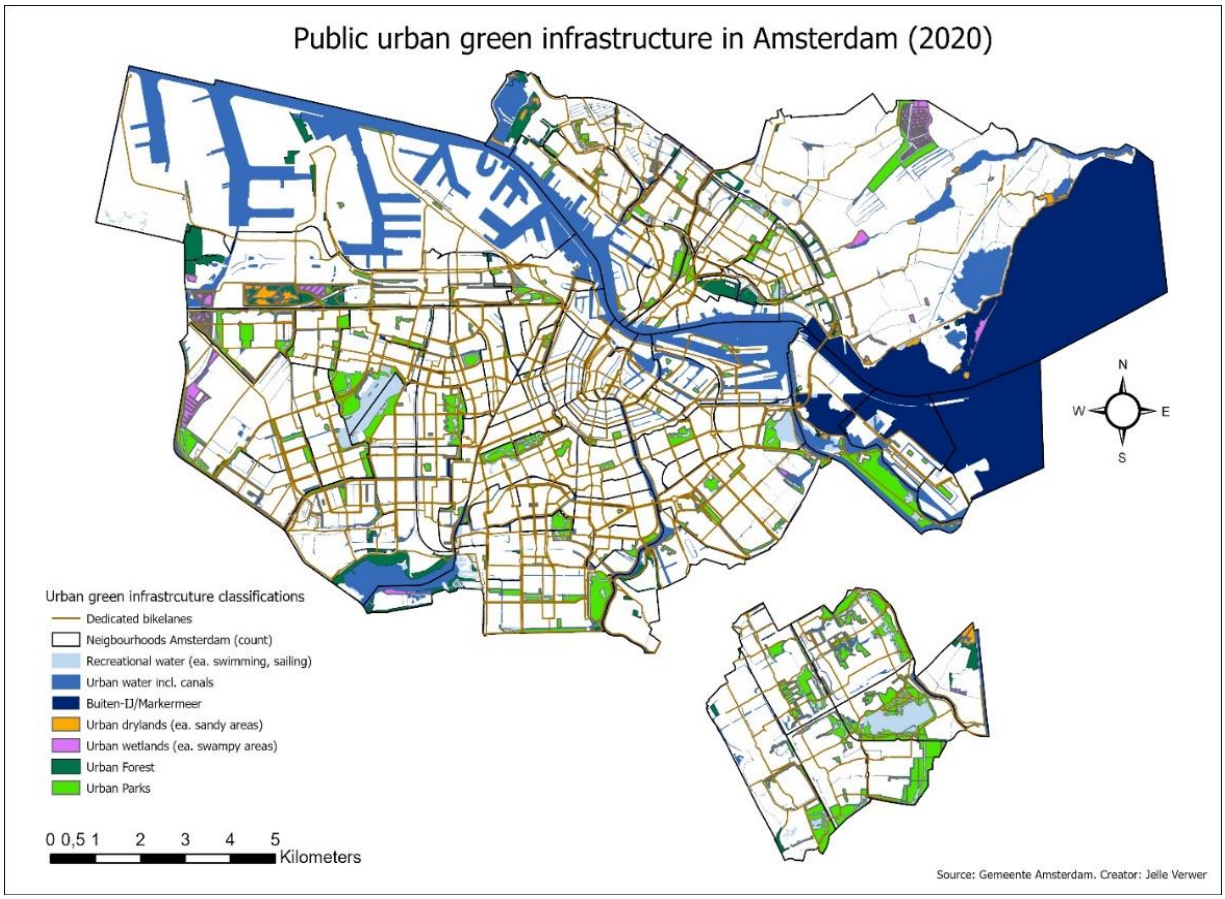


Figure 2-1: *The PUGI within the municipal borders of Amsterdam (Gemeente Amsterdam, 2020).*

2.4 PUGI Policy

Besides elaborate documentation of PUGI and health phenomena the municipality of Amsterdam is pro-actively playing a role in creating a healthier environment for its future citizens. Reason for this may partly lie in the political orientation of the current municipal council and mayor where Groen-Links is the mayor party. The *Green-Left Party* (Dutch: *Groen-Links*), who, as their name suggests, pursue policy that is aimed at increasing PUGI and (social) equality. On Groen-Links' initiative, the municipality is banning/discouraging cars from certain (inner city) parts (Dutch: *Fietsstraten & Auto-luwe straten).* When new roads are laid out bike paths, as a norm, always get more prominent and wider presence, to allow for more and saver usage of non-motorized transport modes (Gemeente Amsterdam, 2017). In addition a € 26.5 million euro investment in the PUGI was made public on the 6th of January 2021 (Gemeente Amsterdam, 2021).

An example of these efforts is the recent changes made to the street design in the dense urban neighbourhood *De Pijp*, in the southern part of the city, where the realisation of a large underground car-parking, made way for some of the street-level parking spaces to be turned into PUGI, bicycle parking and/or widened sidewalks. Simultaneously cars were discouraged from entering the area. Example of the resulting street-design (*impression*) after the transformation is shown in *figure 2-2* to give an impression of the municipal efforts in bettering and increasing the neighbourhood (street-level) PUGI.

Figure 2-2: *The Frans Hals Buurt, part of the neighbourhood De Pijp, recently saw most of its street-level parking be turned into public green infrastructure, furthermore, discouraging car-traffic in favour of non-motorized traffic*

The city of Amsterdam still has many challenges, the intensity of usage of the largescale PUGI (Dutch: *Hoofdgroenstructuur*) is high and the distribution of users over the available green is skew (Zanen et al., 2011). The € 26.5 million euro PUGI-investment is published a new *green* that aims to establish new green infrastructure as well as renovating the existing green infrastructure (Gemeente Amsterdam, 2021). This large investment is made possible by a new green-infrastructure financing structure, where the required funds are covered by a loan that is paid off over a long period of time. This funding-construction is legitimized by the assumption that the PUGI that is invested in will keep its collateral-value for the duration of the loan (Gemeente Amsterdam, 2020). By studying the PUGI in relation to health variables on a neighbourhood

aggregation level in Amsterdam the outcomes can help (future) local policy makers in making and justifying even more investments in Amsterdam-PUGI.

Amsterdam recently (2019) commissioned a study that explored four potential PUGI-development directions in the period 2020-2050 (Paulin et al., 2019). The scenarios in this study are explored because the population forecast of an additional 70.000 people between 2016 and 2025 (803.000 → 900.000), is expected to continue steadily towards 1.000.000 inhabitants in 2030. This long-term population increase within the municipal borders will cause the current PUGI to not be sufficient to accommodate all citizens in using the available public urban green infrastructure for their leisure, social and physical activities. If no ambitious plans will be carried-out, the existing PUGI will be over-extensively used, and possibly avoided by certain citizens in the near future (a trend that is already seen by OIS-Amsterdam in 2018 (OIS, 2018)), see *table 2-1* for this scenario.

Table 2-1: *Tree and park-infrastructure per-person development in the period 2020-2050 (Gemeente Amsterdam, 2020)*

Category↓ Year→	*2020*	*2050*	*Decrease (2020-2050)*
Population Amsterdam	872.380	1.107.823	+27% (absolute)
Trees per capita (nr.)	0.29	0.23	-21.3% (relative)
Parks per capita (m²)	25.2	19.8	-21.3% (relative)

From the 4 scenarios that are introduced (Paulin et al., 2019), the *green neighbourhood scenario* generated the most city-wide benefits according to this research. The *green neighbourhood scenario* focusses on small incremental changes in the urban fabric, for example turning parking spaces into low-vegetation green (see *figure 2-2)*, expanding the presence of green roofs and increasing other low vegetation additions to the existing PUGI. This municipal explorative research however mostly looked at the financial and climate change resilience benefits when comparing the four scenarios; whereas this research will look more closely in how the PUGI impacts health variables. But there are more explorative PUGI-studies in the Amsterdam-area. The Dutch Non-Governmental Organisation (NGO) *de Urbanisten,* who contributed to the 2050 municipal green-vision, specifies some of the 2050 Amsterdam PUGI-interventions and scenarios in another study made in co-operation with the Rijksinstituut voor Volksgezondheid en Milieu (RIVM) (Nijs, 2020). In addition the municipality made a similar 2050 PUGI-vision named *Groenvisie 2050* (Gemeente Amsterdam, 2020). Amsterdam currently ranks well compared to other Dutch cities in the realm of *healthy mobility,* topping the 2020 list of the 20 largest Dutch cities (*figure 2-4, next page)* (Arcadis, 2020, p.30). The exquisite condition of healthy transport possibilities in Amsterdam are confirmed by the 2020 World Cities Report by the United Nations (UN) (UN Habitat, 2020). In the same Arcadis-report that rated Amsterdam as having the *healthiest mobility,* Amsterdam scores the second-worst in *healthy public urban environment (figure 2-3, next page),* only being surpassed by Zaanstad

(Arcadis, 2020, p.32). According to this study, Amsterdam has one of the worst public green availability, public green within viewing distance and tranquil places covered by green (noise, shadow and wind covering) of all large Dutch cities. This worrying shortage of public urban green gives the city of Amsterdam a big disadvantage in becoming a healthy urban environment, one of the sustainable development goals (SDG's) set by the UN (UN Habitat, 2020).

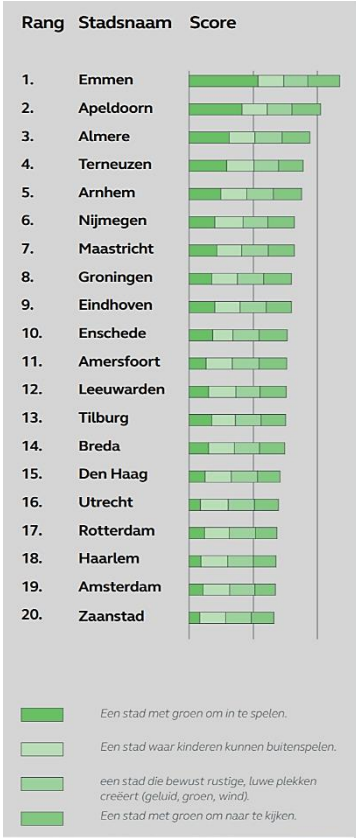


Figure 2-3 Amsterdam is second-worst off the 20 largest cities in the netherlands regarding *healthy public urban environments* (Green infrastructure)

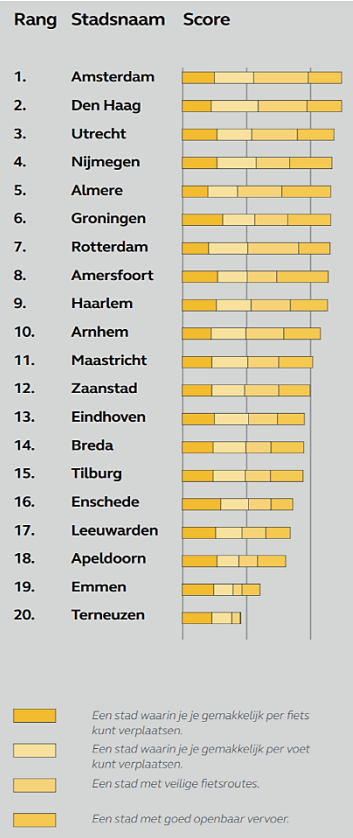


Figure 2-4 At the same time Amsterdam is best off the 20 largest cities in the Netherlands regarding *healthy transport possibilities* (Cycling, Walking, Public transport)

2.5 Conclusion

Concludingly Amsterdam is an heterogeneous urban environment (making comparison between neighbourhoods useful and interesting) with an ambitious city council willing to make PUGI-improvements to positively impact citizen health. In (Nematchoua, Sevin, & Reiter, 2020) the neighbourhood is defined as the preferred scale level of action regarding tackling urban health issues with spatial implementations, therefore this scale level will be used when sufficient data is available for these administrative units. In addition, the future PUGI in Amsterdam is already to some degree explored and can thus be (to some degree) substantiated concerning the potential impact it has on health variables in the year 2050.

3. Theoretical Framework

3.1 Introduction

There is comprehensive research in place that shows specific health-problems that are coming from, or are catalysed by, urbanized environments that lack sufficient PUGI. According to the WHO (World Health Organisation, 2020a) and (Staatsen et al., 2017) the most prominent urban-related health problems are *1. noise and air pollution, 2. the (lack of) greenspace, 3. inadequate transport possibilities, 4. inadequate housing and 5. the lack of space for physical activities*. The direct and indirect relations between these 5 urban health problems and PUGI are elaborated in *table 3-1 (next page).* In (Bosch & Goossen, 2016, p.16) the local (positive) effect of PUGI is explained in a infographic (*appendix 2)*, these health effect show similarities with the WHO urban health problems, but are put in a more local context. (Dinther & Weijers, 2016) in addition show the importance of PUGI as a means to reduce sedentary lifestyles, noise, heat/water stress, air pollution, social seclusion, and stress, from small scale initiatives to large scale interventions.

Substantiation of the relation between more and better quality PUGI and improvements in certain health domains is substantiated in (Bos et al., 2020), for the netherlands in general (nationwide study) and by (Dinther & Weijers, 2016) for an Utrecht neighbourhood. But the potential benefits resulting from more PUGI in relation to health variables is a topic that is also covered by many international studies, for example (Chen & Kan, 2008; Orstad, Szuhany, Tamura, Thorpe, & Jay, 2020; Paulin et al., 2019; Plane & Klodawsky, 2013; Staatsen et al., 2017; Vardoulakis, Salmond, Krafft, & Morawska, 2020). Elaboration of these, and more, relevant studies is shown *table 3-2 (page 18).*

It is worth noting that PUGI is not the only factor contributing to certain health variables, there are several well-known covariables which will be elaborated later on. For example the relation between the socio-economic status (SES) and health (Congdon, 2019). Also the density of the urban fabric (urban core ↔ sprawl/rural) is a proven covariable that impacts health variables (Lopez, 2004). Besides many authors looking into the relation between PUGI and health variables on a local scale, international organizations like the United Nations (UN) put emphasis on this topic. The United Nations and their Un-Habitat 2020 Sustainable Development Goals (SDG) framework confirms that improved PUGI leads to *improvements in urban health* (UN Habitat, 2020).

Table 3-1: Urban related health hazards and the potential impact of PUGI improvements

Urban Health Problems	*PUGI-benefits*
Noise and air pollution	- Roadside green to block traffic nuisance. - Roadside green to filter out $P>M_{10}$ form traffic. - Large scale green for regional air-filtering. - Blue-infrastructure for alternative modes of transport to ease the road-network and decrease nuisance. - Bicycle lanes to increase non-motorized traffic and thus noise/air pollution. - Green roofs for better air-filtering.
A general lack of green space	- PUGI provides lush, shadow-rich, and stress-free environments. - PUGI provides natural environment for physical/social activity. - PUGI provides cooling/shadow (reduce urban heat island effect) . - Green roofs provide extra cooling and air-filtering where street-level green is not possible (dense urban cores).
Inadequate transport possibilities	- Road-side green provide shelter from motorized traffic for non-motorized traffic (increases safety, blocks pollution, and drops shadow). - Create corridors for non-motorized-traffic that connect important nodes within the transport network.
Inadequate housing	- Nearby green to decrease stress among dwellers. - Nearby green increases property value. - Nearby green provides shared outdoor space for small-urban houses with no private green for social and physical activity.
Lack of space for physical activity	- PUGI provides shared activity space (e.g., sport classes, football, picnicking). - PUGI provides personal activity space (running/walking/cycling). - PUGI provides for active ways of transport (walking & cycling).

3.2 PUGI in Relation to Health

A vast amount of theory is in place concerning different PUGI-variables and their relation to citizen health, often in the form of case studies that take place in specific countries or urban-areas. There are a few overarching findings. For example, citizens that are less mobile (in the sense that their lives are constrained mostly to the neighbourhood) are more severely influenced by the neighbourhood PUGI in the city of Toronto (Royal FloraHolland, 2021). This emphasizes the importance of mobility in the green infrastructure; well-planned walking and cycling routes that are surrounded by PUGI to accommodate all-year round easy access to more distant parks and other green-amenities. In general these less-mobile people have a lower socio-economic status (SES), and thus lower income-citizens are more dependent on their neighbourhood-PUGI for physical activity and social life (Plane & Klodawsky, 2013). The UN emphasizes the importance of accessible PUGI as a catalysator for inclusion of minority and suppressed groups in their sustainable development goals (SDG's) and New Urban Agenda (NUA) (UN Habitat, 2020). '*The New Urban Agenda and SDG place an emphasis on inclusive settlements and provide frameworks for unlocking the environmental value of urbanization for all, rather than for a rarefied elite'* (UN Habitat, 2020, p.26).

But not all theory on the relation between PUGI and urban health shows the same relation, in contrast to the previously mentioned findings *Mouratidis* found that deprived urban areas in Oslo often have

more PUGI, but their dwellers are on average less healthy and satisfied with their neighbourhood. This shows that the link between PUGI and health variables is location-dependent (Mouratidis, 2020) and not well generalizable, something that is confirmed by Li et al. and Staatsen et al. (T. Li et al., 2020; Staatsen et al., 2017). In a more broad context Orstad et al. found that PUGI (mainly urban parks) in New York have a positive effect on citizens *mental* health in the direct surroundings (e.g. depression & loneliness) (Orstad et al., 2020). Gunn et al. discovered that in Melbourne, lush and walkable neighbourhoods positively influenced citizen health in many ways (Gunn et al., 2017).

These findings about the (in general) positive influence of PUGI on health variables in urban areas is for the WHO a topic of interest, and therefore they also study these relations in detail (World Health Organisation, 2010). As mentioned, the exact impact that PUGI has on citizen health is case-dependent. Vardoulakis et al. add to this "*Well-planned, sustainable, changes to urban transport, housing, land use, renewable energy generation, and waste management have the potential to lead to improvements in air and water quality and liveability of urban environments, providing multiple benefits including improved public health, reduced inequalities and higher productivity in cities"* (Vardoulakis, Salmond, Krafft, & Morawska, 2020, p.1).

In addition, in Europe, around 350.000 premature deaths are related to particulate matter coming from urban traffic and industry. The incidence of particulate matter is catalysed by a lack of particulate matter retention such as PUGI. Due to PUGI namely, mainly the particulate matter between 2.5 and 10 micrometre in size (PM_{10}), is filtered out (Staatsen et al., 2017; Wesseling et al., 2008). According to a study from the municipality of Amsterdam (Paulin et al., 2019) particulate matter, coming from motorized traffic and industry, also in Amsterdam is one of the biggest urban related health problems, and PUGI (e.g. trees, parks and urban forests) play an important role in retention of these particles and by doing so in creating a healthy environment for urban citizens. Less particulate matter namely leads to less respiratory diseases, cardiovascular diseases, and overall mortality, and thus increase general health-satisfaction (Paulin et al., 2019). In *table 3-2* the most important authors and their findings are elaborated, in *appendix 4* a more elaborative table is present with the research area, study-title and research-methodology included per article and/or author.

Table 3-2: the most important authors and their findings on PUGI-health relationships (see appendix 3 for extended version)

Autor	*Findings*
(Bos et al., 2020)	The effect of green- interventions differ greatly per neighbourhood. The positive impact of green infrastructure is wider then solely the neighbourhood residents. Besides improving social cohesion in general, also the employers, insurance companies and municipality as a whole can benefit from green improvements.
(Dinther & Weijers, 2016)	Green and blue infrastructure produces many ecosystem services; air temperature regulation, air quality improvement, water regulation, noise reduction, mental health improvements, increasing social interaction and physical activity and mitigation of floods and droughts.

(Congdon, 2019)	Low density environments in general lead to less health food in someone's home proximity and thus a higher BMI. People living closer to a public open space (POS) in general have better health in all age groups due to more physical activity. However, the access to POS is related to ethnicity and income, thus these are important covariables. Besides the build environment, the social environment in a neighbourhood (measured by social capital) is also explaining variance in obesity.
(Doiron et al., 2020)	The least deprived areas were in general more walkable, had lesser pollution and more green-infrastructure. This result suggests at least in Canada, environmental inequality is prevalent. For these variables, in general socio economic status (SES), built environment and public health are closely connected.
(Fan et al., 2011)	Neighbourhood green has several distinct roles in reducing stress; parks mitigate stress indirectly by fostering social support. Park/green size has an overall more positive effect then the overall neighbourhood green level. It is important green spaces in urban areas accommodate socialization opportunities.
(Gunn et al., 2017)	Neighbourhood activity centres (clusters of amenities like shops, barbers, and social life) were analysed on their walkability in regard to their surrounding customers. Outcome was that among other things the connectiveness of the street network, the diversity of the NAC and the residential density impacted the % of residents that went to them walking positively when being high.
(Holtan et al., 2015)	There is in this study a positive relationship between the neighbourhood tree canopy cover and the social capital of individuals.
(Jennings & Bamkole, 2019)	Social cohesion Is associated with increased physical and physiological health benefits. The presence of urban green infrastructure can increase social cohesion and thus enhance health and well-being. Especially the social interaction within urban green infrastructure is enhancing and catalysing social cohesion, social capital, and healthy behaviour.
(Royal FloraHolland, 2021)	Green infrastructure leads to health benefits, e.g., less ADHD prevalence, less anxiety disorders, less depression, less cardi-vascular diseases, less heath in summer, less stress, more outdoor physical activity, less water nuisance/flooding, less headaches, and better concentration.
(Klompmaker et al., 2018)	Nearest green infrastructure is not associated with obesity or physical activity according to the results. The overall NDVI is however correlated with obesity and physical activity. The associations where stronger in less-urban areas and with smaller buffers (importance of proximity).
(Kumar et al., 2019)	Rapid urbanization turns cities into concrete landscapes. While inclusion of PUGI may support health, socioeconomic support, and environmental improvements. Strategic placement of PUGI is however not related to reduced air-pollution. Thus urban-greening is not a significant measure to mitigate air-pollution.
(F. Li, Zhou, & Lan, 2021)	The urban-form (level of fragmentation, sparseness, greenery etc.) can have an impact on the air quality. This relation is substantiated on different scale-levels using a regression-model and the different urban forms impacted air quality more when using decreases scale-levels. IN general poly-centric scattered and forested urban forms are resulting in the best air-quality.
(Lopez, 2004)	Sprawl increases the risk of being overweight or obese significantly. For every 1-point (1-100 scale) rise in sprawl overweight population increased .2% and obese population with .5%. Concluding sprawl is one of the drivers of increased BMI.
(Mowafi et al., 2012)	Studies on the PUGI-health relation have been conducted in developed countries generally, this study studies Cairo neighbourhoods. The study found that PUGI is not related to BMI in Cairo, a conclusion that is contradicting the general findings in developed cities and countries.

(Nørregaard, Gram, Vigelsoe, Wiuff, & Birk, 2014)	The distance to green-spaces in relation to physical activity and BMI is looked into. Living more then 1-kilometre from PUGI does indeed decrease the odds of physical activity. Also, obesity was more prevalent among people living more then 1-kiloemtre from green-spaces.
(Norton et al., 2015)	Global warming in combination with urban developments and urbanisation is potentially deadly due to the resulting weather-extremes. The quantification of the cooling benefits of PUGI is elaborated and a framework is made that prioritises types of PUGI based upon their colling abilities.
(Silvennoinen, 2017)	This study discusses to what extent people's decision-making in the choice of transportation mode is influenced by urban design factors. The results show that residential location and daily travel distance influenced the inhabitants' purpose and reason to cycle, but were not significant when cycling was not chosen as a transportation alternative (quoted from abstract on page 5)
(Vardoulakis et al., 2020)	The SDG's aim to increase sustainable lifestyles by 2030, due to the majority of the people living in urban-environments, implanting urban interventions that will improve health, wellbeing and sustainability in the built environment is essential. Promoting climate-sensitive urban planning and policy is an important means to improve health; think zero-carbon urban housing, active transport, and heat and flood protection.

3.3 Covariables

3.3.1 Introduction

In addition to some studies showing a somewhat isolated relation between PUGI and health, many theories are supporting the concept that several covariables are of importance. When crucial covariables are overlooked, the regression-model misses some crucial information to predict the relation correctly, reducing the absolute and relative model-fit. Therefore, covariables are thought about in advance. Covariables that are repeatedly present in literature are socio-economic factors of a neighbourhood and the level of urbanisation, and thus density of the urban fabric. These covariables will be elaborated in the upcoming paragraphs.

3.3.2 Socio-Economic Status

Repeatedly mentioned in the theory is that the addition of PUGI within an area improves quality of life (e.g. social life & work productivity) and health (e.g. physical and mental) among all urban dwellers in general, but does so *more* for people living in neighbourhoods with a higher SES (Bos et al., 2020; Congdon, 2019; Doiron et al., 2020; Gunn et al., 2017). Doiron et al: *'Results from our analyses suggest that such co-benefits of greenness exposure might be disproportionately experienced across different socio-economic status (SES) levels.'* (Doiron et al., 2020, p. 7). These *co-benefits* that are mentioned by Doiron et al. that impacted by the PUGI are in this article diabetes, physical activity, birth problems and general mortality (Doiron et al., 2020). In short, high SES-level neighbourhoods have *relative* higher benefits then low SES-level neighbourhoods concerning local PUGI. Congdon supports these findings and claims: '*Both access and quality of parks may depend on the socio-economic level of neighbourhoods*' (Congdon, 2019, p. 2). One of the causes may be that high SES-neighbourhoods have an overall higher *quality* of their local PUGI. For example, free-to-use gym-equipment, dedicated jogging-routes, better green/blue maintenance and/or better catering industry

presence to increase the locational quality and safety.

The way people move around and transport themselves, and thus to what degree they undertake a certain mode of transport, that incorporates physical activity, is dependent on their SES (Gunn et al., 2017; Westfall & Villa, 2001)(Gunn et al., 2017). Increased walkability within an area for example is in Melbourne (Gunn et al., 2017) connected with increased average personal income of an area, and thus higher SES. But even without socio-economic characteristics considered, it can be expected that the urban environment, from which PUGI is a part, is an important determinator of health. Congdon adds to this: '*Environmental factors are important influences on changing levels of physical activity and on changing dietary behaviours throughout the life course'* (Congdon, 2019, p. 1).

There is however another relation between SES and PUGI, as the UN points out. Increased urban greening can namely also *displace* lower SES people. *'Recent assessments of urban greening initiatives show that while they have resulted in positive environmental outcomes (increase in green space, reduction of pollution), they have also been associated with the displacement of low-income residents.'* (UN Habitat, 2020, p. 19). Resulting in a complex situation were often increased PUGI, meant to help the least healthy population (low-SES households), also negatively impacts them due to the green often being built on locations were current low-SES citizens life. This phenomenon is known as *green-gentrification* and is best to be avoided to maximise the positive impact of PUGI on *all* inhabitants.

3.3.3 Urban Density

The degree of urbanisation is a covariable that explains health of urban dwellers to some statistical significance in existing research. In (Congdon, 2019), urbanity is seen as one of the covariables for determining health-variables and in (Lopez, 2004) a positive relation between urban density and obesity is found, where less dense areas have higher obesity prevalence. High density urban neighbourhoods are characterized by walkability, amenity proximity (due to mixed-use) and car-independency, and thus less particulate matter and more usage of alternative, non-motorized modes of transport. Urban dwellers in high density urban areas have in general a stronger focus on the local PUGI, whereas rural/sprawl dwellers live in low-density environments, functionally segregated, car dependent and with less focus on local (neighbourhood) characteristics (Congdon, 2019).

Due to this difference in physical environment and hereof coming transport and social-life discrepancies, urban and rural/sprawled environments are hardly comparable, therefore non-urban and rural neighbourhoods within the municipality of Amsterdam are excluded from the research-scope. Gunn et al. underlines the concept that dense and highly urban areas provide for better active ways of moving around: *'A number of built environment features are consistently shown to facilitate transport walking around residential homes, which are the origins of many walking trips. These include: highly connected streets, high population density, mixed land use and good access to destinations and transit, and sidewalk provision'* (Gunn et al., 2017, p. 2).

There are several methods to group neighbourhoods and cities on urbanity, for example the CBS address density (addresses per km²), the official urbanity measure of the Dutch bureau of statistics, but also the Floor Space Index (FSI) can be a good predictor to the density and urbanity of an area (Milgen, 2016). In addition to the address-density, the official urban-classification system in the Netherlands, internationally seen, the FSI is most used to measure the level of urbanity of an area (Harbers, Spoon, Amsterdam, & Schuit, 2019). The address-density is available on neighbourhood-level, while the FSI is available on street-level in Amsterdam.

The FSI, sometimes referred to as the Floor-area ratio (FAR) is a measure that is often used by zoning laws to limit and push planners and developers to a certain building pattern (Patel, 2013). The FSI has the advantage that it allows for flexibility with the developer as one can build higher but reduce the footprint or build low rises and widen the footprint while maintaining the same FSI. The FSI for some major world-cities is shown in *table 3-3*, the table shows how Amsterdam compares globally. Amsterdam has a relatively low FSI. In the table only the core-city boundaries are shown here and not the edge-cities and sprawled areas. This leads to some north-American cities ranking higher than you would expect.

Table 3-3: the FSI compared for some world-cities

City	*FSI*
Sao Paulo	1
Mumbai	1.33
Amsterdam	1.9
Venice	2.4
Paris	3
Shanghai	8
Vancouver	9
San Francisco	9
Hong Kong	12
New York	15
Tokyo	20
Singapore	25

4. Methodology

4.1 Introduction

The methods consist of 4 parts, where each part covers a research question. Methods for reviewing the existing literature (RQ1), methods for findings and pre-processing datasets (RQ2), methods for analysing the datasets using a multiple regression (RQ3) and methods for reviewing the 2050-PUGI vision (RQ4). The methods will start with an overview of the research area and the spatial analysis pitfalls, related to a multiple regression analysis. All of the analysis-methods use software that is widely used and recognized for this purpose, namely ArcGisPro (free alternative: QGIS), GeoDa (already open-source and free), Excel (free alternative: OpenOffice/Google suite) and FME (free alternative: GeoKettle/Hale studio). Using these free and/or open-source alternatives, and using available open-data, this can be reproduced without any financial resources. The research methods per research question are elaborated in the upcoming sub-chapters. To start off with a table overview of the methods per RQ's analysis steps is presented for clarification (*table 4-1*).

Table 4-1: Analysis m*ethods per research question*

Research Question	*Analysis steps (in brackets the tools/platforms that are used)*
RQ1: What theoretical relationships are found between PUGI-variables and health-variables?	- Analyse existing research on urban health, urban green infrastructure, and the relation between these subjects (literature review). - Summarize and look critically at the data that these researchers use, their methods and findings/conclusions (literature review). - Filter out what theory, methods and finding are relevant for this study for next RQ (literature review).
RQ2: What PUGI-variables, health-variables and co-variables are openly accessible (on a neighbourhood level in Amsterdam)?	- Gather the relevant datasets (independent variables, dependent variables and covariables) that are accessible via open geodata-portals (data review). - Filter and select critically, e.g., what are their temporal validity, is the source trustworthy and/or is the spatial resolution sufficient? (data review). - Pre-process 1: Transform them into workable formats (FME Workbench; Excel). - Pre-process 2: Eliminate unwanted features outside of scope (ArcGisPro; Excel). - Pre-process 3: Aggregate the spatial data to the required level: neighbourhoods (ArcGisPro). - Pre-process 4: Join all non-spatial characteristics to the neighbourhood aggregate (ArcGisPro).

RQ3: What statistical relationships are found between PUGI and health variables among neighbourhoods in Amsterdam?	- Normalize or tesselate the neighbourhood data in order to overcome the MAUP (ArcGisPro). - Create a neighbourhood matrix in order to cope with proximity and neighbour influence (GeoDa). - Run the OLS regression with the PUGI, health and co-variables (BMI to start off with) (GeoDa). - Depending on the results choose whether to use an OLS, Spatial LAG or Spatial Error model.(*see appendix 1*). - Analyse of the regression results; check for heteroscedasticity (check residuals), check for clustering, spatial autocorrelation (global Moran's I), statistical significance (P-values) and ecological fallacy (covariables) (GeoDa). - Visualise spatial regression results (ArcGisPro; GeoDa). - Create maps and other statistical visualisations (e.g., histograms, boxplots, and LISA-maps) in order to interpret the spatial findings. (GeoDa; ArcGisPro). - Interpret and compare the results with the existing theory in mind (Literature review; GeoDa). - Look into the spatial pitfalls and see if they pose any problem with the reliability of the outcome (Literature review; GeoDa). - Validate the model outcomes, statistically or theoretically 9GeoDa; Literature review).
RQ4: How can future public green infrastructure impact these relationships between PUGI-variables and health-variables?	- Create an overview of all urban green infrastructure visions of Amsterdam and make a list of all significant measures. (Literature review). - Use the outcomes of the 2020-regression results to see what interventions are most viable and useful in relation to health improvements. - Analyse the green infrastructure 2050-vision using the known 2020 relations.

4.2 Literature Review

The literature review consists of creating an overview of all existing PUGI-related literature; from which a large part is already covered in the theoretical framework. Grouping of the existing body of literature by author and concept is used as a means to create an overview of the theory. The literature review helps establish a framework within which the data-collection and analysis can take place. Creating an overview of widely studied PUGI-health relations and theories that are used ensures the data redundancy is as little as possible and the scope of the research stays as narrow as possible.

4.3 Spatial Scope

The spatial extent of the research are the Amsterdam municipal borders. This extent is chosen due the interoperability and accessibility of datasets on this

Table 4-2: relevant scale-levels for spatial interventions in the urban green infrastructure (Staatsen et al., 2017, p39)

Strategic ←						→ Local
Region	City	Town	District	Neighbourhood	Street	Block
Land-Use patterns						
Transport						
Green Infrastructure						
				Urban Design		

level. The neighbourhood scale level is within the municipal extent the chosen aggregation level. Partly due to the data-availability on this level, but also based upon research from the World Health Organization (WHO) (Staatsen et al., 2017). It is mentioned by the WHO that the neighbourhood scale level is most relevant for urban interventions as it is the only scale level that relates to all relevant urban planning components, this is shown in *table 4-2 (previous page).* This is important for finding appropriate datasets on the topics that are deemed relevant according to the literature review, as not all data is available on this fine-grained aggregation level.

4.4 Data Collection

Spatial Data portals form the gateway to the required datasets. The Netherlands has many high-quality open-data portals. These portals are mostly government-funded, and the datasets are often collaborations between many governmental organisations, research institutes and some public-private parties. Many portals have some overlap in their offering, for example data.overheid and PDOK cover some of the same popular datasets. The national georegister, within the Netherlands, functions as the metadata-portal for alle geographical data, therefore this is a great portal to start off with as there are over 7.713 geographical datasets available here. The georegister does not host any data itself, often you get redirected to the data-supplier, for example PDOK, Kadaster, Rijkswaterstaat or municipal/provincial-data portals. *Table 4-3* gives an overview of the portals that are used. More elaborate version of this table, with relevant URL's, is available in *appendix 7.*

Table 4-3: the data-portals that will be used for the analysis

Portal	*Scale*	*Funding*	*Input*
Data.overheid	National	Ministerie van binnenlands zaken en Koninkrijksrelaties	>180 governmental organisations
CBS-statline	National	Ministerie van Economische Zaken en Klimaat	CBS (independent organisation)
Data. Amsterdam	Amsterdam	Municipality of Amsterdam	Municipal input (OIS) + external partners
RIVM-statline	National	Ministerie van volksgezondheid, Welzijn en Sport	The responsible ministry
PDOK/Nationaal Georegister	National	Ministerie van infrastructuur en waterstaat, Ministerie binnenlandse zaken en koninkrijksrelaties, Ministerie van economische zaken en klimaat, Rijkswaterstaat, Geonovum	>450 organisations
ArcGIS University of Groningen geo-portal	National	University of Groningen (RUG)	University of Groningen

4.5 Data Preparation

The datasets shown in *appendix 3* covering the required dependent and independent variables are viewed, inspected, and altered using GIS-software; the software of choice is depending on the steps required to transform the raw-data into usable data. Knowing from experience and literature, FME workbench, Excel, ArcGisPro and/or QGIS will be used as they offer a wide range of data-conversion, viewing, inspection and alteration options. FME-workbench will be used mainly for data-format conversion and will only be used before importing the data into GIS software. Excel is used to inspect and alter the tables containing the population statistics and turning them into a CSV before importing them into ArcGisPro.

Important to know is what type of data is used; Nominal, Ordinal/Categorical, Interval or Ratio data (David O'Sullivan; David J. Unwin, 2010), see *figure 4-1* for clarification. Categorial-data has no order, and the values given to the entries are purely to distinguish them from each-other, like the CBS land-classification. Here a CBS70-classification is not better/higher than a CBS35-classification. Nominal data has to be mutually exclusive and inclusive at the same time, meaning an area can only belong to 1 category and no objects are outside of any category. Nominal data is hard to use as input in a regression model or any type of spatial statistical analysis, and therefore it has to be quantified to a higher scale like ordinal, Interval or Ratio (or a binary dummy variable). This is for example done by measuring the amount of a certain CBS-category surface area per neighbourhood, or neighbourhood buffer.

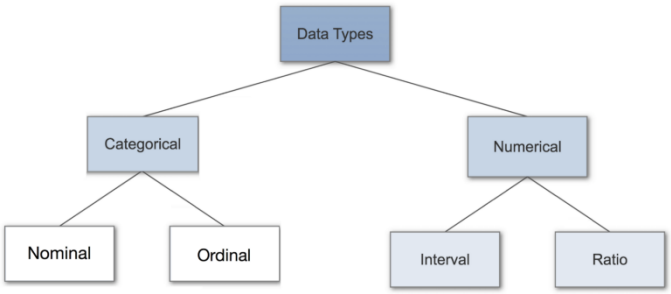


Figure 4-1: the 4 most common data-type distinctions

Non-numerical ordinal data is not used in this study. The CBS-land classification data would be ordinal if the PUGI-land classifications would be ordered in a way that better green infrastructure would get a higher CBS-classification. Finally, interval and ratio data-types are due to their usefulness in quantitative analysis seen as a higher order data types as they allow for immediate statistical and mathematical use. Interval data however has no absolute 0 and therefore can only be used to measure difference but not relative or absolute magnitudes (David O'Sullivan; David J. Unwin, 2010). Often temperature is used as an example of interval data, as within a temperature-scale the difference of 10°C is the same everywhere, but a temperature of 20 degrees is not twice as high as a temperature of 10 degrees. This is inherent to scales where there is no absolute 0. This brings us to the final category; Ratio-data. This category is for example seen in the percentual obesity data, where a 10% obesity rate is exactly twice that of a 5% obesity rate. And thus, comparison is straight forward. Only Nominal and Ratio data are used in this study, there are no datasets with non-numerical ordering (ordinal-data) or numerical data with no absolute-0 (interval-data).

4.6 Data Filtering

The datasets are filtered on the geographical and analytical scope: removing features that are outside of the research area and scope. This is required because some datasets also include input from neighbouring municipalities such as Amstelveen of Ouderkerk aan den Amstel. These are filtered out, only showing the features within the 99 neighbourhoods of Amsterdam. The address-density urbanity classification is hereafter used to filter out the non-urban neighbourhoods. These two filtering methods make sure no redundant geographical areas are included in the dataset. The datasets that contain data outside of the analytical scope, and have no affiliation with PUGI and or health, or other importance for the analysis, are also excluded from the dataset to reduce cluttering.

4.7 Data Aggregation

The useful PUGI-variables, health-variables and covariables are aggregated to a corresponding administrative aggregation units (neighbourhoods). This is due to data-availability on this aggregation level, but this aggregation-level is also useful for easier policy implementation, were the neighbourhood level is the most appropriate level according the WHO to implement green infrastructure measures (Staatsen et al., 2017). Aggregating to this level ensures a weights matrix can be calculated for all neighbourhoods to show what their geographical neighbours are, a necessity for spatial-autocorrelation diagnostics and hereof coming interpretation. Neighbourhoods namely have interdependencies with other *neighbouring* neighbourhoods, for example due to the impact of Tobler's-law (Tobler, 1970).

4.8 Multiple-regression analysis

4.8.1 Model Types and Indicators

Modelling in GeoDa can be done using 3 model-types: an ordinary least squares-model (OLS), a spatial-lag model and a spatial-error model. The first model-type is seen as a classical regression model with a constant-coefficient and variable-coefficients, together modelling the 'ideal regression-line, where the squared residuals are lowest. This type of model however has little incorporation of the spatial dependence between the variables (spatial autocorrelation; SA). An OLS-model on spatially autocorrelated data leads to large discrepancies in the residuals between the observations (neighbourhoods). A spatial-lag model (spatial autocorrelation model; SAR) would be more fitting in this case, where a lag-coefficient is added to the standard constant and variables coefficients. This lag-coefficients uses the neighbourhood structure to incorporate a value that weights in the neighbouring values to account for the interaction between observations. Finally, a spatial-error model is an option. This model-type is often used when SA is a nuisance rather than a substantive part of the research. The spatial-error model uses the errors in a model to account for their dependence, compared to the actual values in the spatial-lag model. The spatial-error model is best

used when the SA is an nuisance to the model performance, whereas the spatial-lag model is more used for substantive modelling purposes.

Table 4-4: important model diagnostics and their use from the literature review

Phenomena	*Indicator*	*Explanation*	*Interpretation*
Variance explanation/ absolute model fit	R-squared (R^2)	This indicator lies between 0 and 1. A higher score means the explanatory variables explain a large part of the variance of the dependent variables.	High value (e.g., >0.8) means a strong model fit, 80% of the variance is explained in this case.
Variable co-efficient	Numeric value (positive or negative)	The coefficient-value describes the regression model-line and in addition shows what direction and what strength the relationship between explanatory and dependent variable is (positive, neutral, or negative)	Coefficient-value can be below or above zero (positive or negative correlation). Having a p-value below .05 is needed for significance.
Variable probability	p-value	The p-value shows what the likelihood is that the variable, or other regression test-statistic is statistically significant.	Lower is more significant, a value below .05 in general is significant.
Relative model quality	AIC, Log-likelihood & Schwarz-criterion	The AIC-value shows a trade-off between the model-fit and the model-simplicity. The value of the log-likelihood is negative, the AIC and Schwarz-criterion positive.	Values closer to 0 mean model-fit is relative better (so lower AIC and Schwarz-values in combination with a higher log-likelihood.
Spatial autocorrelation of the residuals	Moran's I	The residuals of every aggregation unit are used to measure if they are spatially autocorrelated, meaning underpredicted and over predicted values are statistically significantly correlated	Positive numbers indicate positive auto-correlation and vice versa, statistical significance is required in order to accept the presence of spatial autocorrelation
Multicollinearity	Multicollinearity test	Shows when explanatory variables are correlated with each other and the dependent variables and thus implies redundancy in the variables.	A value below 30 indicates multicollinearity is acceptable
Model choice	Lagrange multiplier test	Shows which alternative model can be used from the OLS-baseline (Error and/or Lag model)	You can best choose the model that has the lowest p-value with the Lagrange multiplier test.
Heteroskedasticity	Breusch-Pagan test and/or Koenker-Bassett test	When these tests are significant this means the standard errors from the OLS are incorrectly calculated, this happens for example when within high-green areas there is more variation in obesity then in low-green areas.	Using robust standard errors/do nothing if sample-size is large enough

4.8.2 Spatial Autocorrelation

SA is a direct result from the general consensus that samples from nearby locations are more similar than samples remote from one another. A phenomena first described in detail by Waldo R. Tobler in a Detroit case study in 1970 (Tobler, 1970). In short this phenomena ensures spatial data in general violates the concept of sample randomness that normally is required for statistical-analysis (David O'Sullivan; David J.

Unwin, 2010). Spatial data, on the other hand, especially has our interest due to the data being not random over space, and thus the SA imposes a contradiction. SA introduces redundancy in the data samples as we can assume nearby samples are more similar. A commonly used diagnostic to measure the amount of SA is the global Moran's I. SA diagnostics can, when being significant, also indicate a standard OLS-model is not most fitting, and instead a spatial-error model or spatial-lag model should be used. These models can help account for the SA in the residuals to improve model-fit. SA can be negative, zero or positive. Positive SA is the most common as said, due to the general assumption that nearby samples are more similar. Zero SA implies spatial randomness in the data samples. Negative SA on the other hand implies that the more far away a sample is the more similar it is. This is a uncommon phenomenon. SA cannot be 'solved' as it is inherent to the data that is analysed but the diagnostics can warn us from the strength of the effect and take measures accordingly to some degree (model-choice for example, or when the model is substantive SA-interpretation is a goal in itself).

SA can be a nuisance to models that aim to obtain proper statistical inference. However, SA is less of a problem for models that try to discover a spatial relationship/interaction, in this case SA is considered as substantive. Finding out that there is in fact SA has then be taken in consideration with the interpretation of the model. The SA is then less a model-nuisance and more of a statistical finding. The spatial-lag model and/or spatial-error model are potential models that help interpret the SA by using neighbourhood structures to create an extra coefficient. The model performance (R-squared) and relative model fit are then in theory improved. Deciding on what model is expected to have best fit with the used data can be determined using the SA-statistics in the OLS-model and using the model-decision-tree *(Appendix 1).*

4.8.3 Modifiable Area Unit Problem

The modifiable area unit problem (MAUP) occurs when data is aggregated to a non-natural level. Often the most common governmental and statistical units like neighbourhoods, municipalities, or states, are not per se the aggregation units in which a spatial phenomenon is bounded by. In other words: these aggregation levels are arbitrary in relation to the phenomena. This discrepancy is known as the modifiable area-unit problem (David O'Sullivan; David J. Unwin, 2010). The effect is for example seen when a aggregation level is changed and the resulting spatial relations from the phenomena under interest change with this. In general, a size increases in the aggregation unit (e.g., neighbourhood → municipality), goes together with an increase in the strength of the regression relationships (R^2). This is due to the fact that larger geographical units often present samples that have their mean lie closer to the population-average (and thus to the regression line) and the effect of outliers is muffled. A topical example of the effect MAUP has on people's real life is the United States census tract grouping, often referred to as Gerrymandering, named after a 18th century Massachusetts governor. This sophisticated grouping of census tracts in combination with a nation-wide winner takes it all voting system, leads to a situation where one can win elections in the United States with less than 50% of the popular vote. The MAUP is hard to account for due to the inherent aggregation of most

data to administrative regions. What can be done however, when individual data is not available, is discussing what the impact of the aggregation unit is on the outcome of the analysis.

4.8.4 Ecological Fallacy

Ecological fallacy is a problem that occurs when the level of aggregation is not the same as the level of the phenomena studied (David O'Sullivan; David J. Unwin, 2010). It is therefore closely related to the MAUP but is part of a more comprehensive statistical pitfall. The ecological fallacy implies that when a statistical relation is seen at, say neighbourhood level, we can never assume that this will also be the case at municipal level, or any other aggregation level. The mechanisms that drive the relation may change according to aggregation level. When we see a relation between PUGI and certain health-variables on a neighbourhood level, all we can conclude is that this relation is present at this exact level. We cannot assume there is any relation between PUGI and health variables on a street or municipal level. Other covariables and underlying trends may cause relations to be different that are aggregation-level dependent. Therefore, any existing literature on PUGI in relation to health that did not aggregate to neighbourhoods in Amsterdam is a priori not generalizable to this case; only *assumptions* about the expected relations can be made.

4.8.5 Edge Effects

Edge effects are caused by the demarcation of a study area. Within spatial statistics it is assumed that neighbouring samples and entities influence each other, but at the edge of a research area this is not possible as a part of the neighbouring samples, if available at all, are off the grid *(figure 4-2)*. Edge effects can be prevented by demarcating the research area broader than the actual area of interest. In this study the research area is the municipality of Amsterdam and data from the surrounding municipalities often has discrepancies with the Amsterdam data and therefore the edge effects are hard to account for in this case.

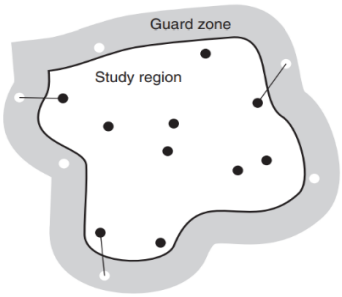


Figure 4-2: Edge effects visualized.

4.8.6 Validation

Validation of the multiple regression model on the topic of interest can be established in several manners. First of all, the regression model can have a high absolute model fit (R^2), giving a strong reason to believe that the model has a great fit and the outcomes are robust. In addition, the existing theory could help validation; if a certain regression-outcome is substantiated by a vast body of theory; this would make the outcomes very likely to be correct in this instance too. Another validation-method that is often used is the *leave-one-out* method, were one at a time the independent variables are eft out to see what this does to the determination, probability and other regression-statistics (Makhatkov, 2019).

4.8.7 Spatial Patterns

It is important to examine the PUGI and health data in order to find if the data is spatially random or has a significant clustering. If the data is spatially random (the null-hypothesis) a classic OLS-model is most fitting. When SA is present, a spatial model (lag and/or error) has supposedly a better fit due to inclusion of an extra matrix-dependent coefficient (respectively a lag-coefficient: βW or a LAMBDA-coefficient: Λ). Due to the fact that visual interpretation of patterns and clustering is subjective (often overestimates correlation, as people are programmed to see patterns) a statistical test has to be used. Often Z-scores and P-values are used for this purpose. The lower the P-value the lower the chance a spatial distribution is random; a 0.01 P-value means the chance a variable is spatially random is below 1% (0.01*100), a 0.05 P-value predicts a 5% chance that the pattern is random, and so on. Z-scores add to this that they can be positive and negative and thus can imply specific cold and hot spots. Therefore, the combination of P-value and Z-score can show what chance there is that a pattern is random. The scores can in addition show if under or over-estimation is present and to what degree this happens in certain clusters, detecting hot and cold-spots.

4.8.8 Standardization

In order to make the data better comparable between all neighbourhoods the variables that contain absolute values that have a relation to neighbourhood-area (and thus tend to be higher in larger areas) need to be aggregated to comparable scale-levels (tessellation) or undergo standardization values. Standardization as a result reduces the value-discrepancy between neighbourhoods with different surface areas. Standardizing the data means the values of a variable are re-calculated in such a way the average of all observations is 0 with a standard deviation of 1, this also called the Z-score. The formula for standardization is shown here (1).

$$X\_standardized = (X - X\_Average)/X\_\ standard\_deviation. \quad (1)$$

4.8.9 Matrices

To incorporate the interdependence of nearby features on a central feature, matrices can show interaction, adjacency and neighbourhoods in a convenient way (David O'Sullivan; David J. Unwin, 2010). Neighbourhood matrices show and help modelling what neighbouring features impact each other, and also to what degree they do. Often giving nearby samples and features a larger impact and stronger relation that translates for example to the lag-coefficient in a spatial-lag model. Matrices are often shown using brackets; in the previous example (*figure 4-3)* the Euclidean distance between 6 features is shown (David O'Sullivan; David J. Unwin, 2010, p.47). The top left to bottom right diagonal line is zero; these values correspond to the same observations (e.g., the distance between observation A and observation A is 0).

$$\mathbf{D} = \begin{bmatrix} 0 & 66 & 68 & 68 & 24 & 41 \\ 66 & 0 & 51 & 110 & 99 & 101 \\ 68 & 51 & 0 & 67 & 91 & 116 \\ 68 & 110 & 67 & 0 & 60 & 108 \\ 24 & 99 & 91 & 60 & 0 & 45 \\ 41 & 101 & 116 & 108 & 45 & 0 \end{bmatrix}$$

Figure 4-3: Distance matrix showing the distance between 6 imaginary objects

A contiguity-matrix can also be used to show solely adjacency between samples and polygons (*figure 4-5)*, this will take the form of a binary table, were for example all observations within 50 meter of an observation are shown as 1 and all further-away observation with 0 (David O'Sullivan; David J. Unwin, 2010, p.48). This adjacency matrix can also be set to include only the K-nearest neighbours (e.g., K5), or any other rule to include a particular set of adjacent observations into the neighbourhood. There are several options for neighbourhood matrices. 3 main possibilities are considered. 1 possibility is to look at a distance (In a distance unit like meters) from the centre of a selected feature. For example, all samples within 1 kilometre of the central neighbourhood get a positive value (1) in the matrix. Another possibility is to look at the contiguity, the bordering of edges of nearby features, where the features that touch edges (rooks) or corners and edges (queens) are included in the matrix. This can then be done for the 1st degree neighbours or the 2nd degree and higher neighbours (the neighbour's neighbours). This type of neighbourhood matrix is required for a spatial-lag model and spatial-error model as they require symmetric weights. Thirdly the K-nearest neighbour method is a possibility, this one looks at the *K(N)* neighbours. The absolute size of the neighbourhood thus is dependent on the density, position, and size of the observations in a certain spatial dataset. Finally, a neighbourhood can be made using time, this is often done when a transport-mode is present that is predominantly used, for example a car. Then all features within a 30-minute car-ride are seen as part of the neighbourhood. For this research however the multi-transport oriented city of Amsterdam makes this type of neighbourhood -matrix hard to substantiate. Different modes of transport are used by different neighbourhoods and also combinations of transports are common in Amsterdam, for example taking the bike with you in the metro or walking to a distant car-park before using the car.

$$\mathbf{A}_{d\leq 50} = \begin{bmatrix} * & 0 & 0 & 0 & 1 & 1 \\ 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 \\ 1 & 0 & 0 & 0 & * & 1 \\ 1 & 0 & 0 & 0 & 1 & * \end{bmatrix}$$

Figure 4-4: Adjacency matrix showing all observations within 50 meter of observation A

Advantage of the K-nearest neighbour method is that the island polygons (topological islands, not actual island) that are present in the dataset get neighbours assigned, whereas with a contiguity matrix these polygons would remain standalone features with no relations in their matrix. The distance-method will calculate an average distance for all neighbourhoods to their neighbours in Amsterdam, this distance will be expected to be too high due to the minimal distance between Amsterdam south-east and the centre part of the city of 1.498 meters. This minimal neighbourhood distance will be too high for mainly the small and dense inner-city neighbourhoods.

4.9 The 2050 PUGI policy

4.9.1 Introduction

The 2050 PUGI is analysed, based upon the *Groenvisie 2050* (Gemeente Amsterdam, 2020). The 2050 PUGI-intervention are analysed based upon the results from the 2020 PUGI multiple regression analysis. The 2050 PUGI-policy measures are compared to the PUGI-health relationships that are found in the multiple-regression to find out if the 2050 PUGI is expected to have an positive effect on the health of citizens. The 2050 PUGI-policy (*Groenvisie 2050)* is at mostly containing non-visualized policy-measures. However, the NGO *de Urbanisten,* in consultation with the RIVM, visualized some PUGI-scenarios themselves. 2 of these thematic maps are shown in *figure 4-5* and *figure 4-6* to give an impression of the different scenarios that are considered. The figures correspond respectively to the local-oriented *green neighbourhood scenario* and the regional-orientated *metropolitan landscape scenario* (Paulin et al., 2019). However, these RIVM/Urbanisten scenarios are mainly aimed at quantifying the (monetary) benefits of the strategies, whereas the *Groenvisie 2050* contains the actual policy implementations, and therefore is more useful for analysis.

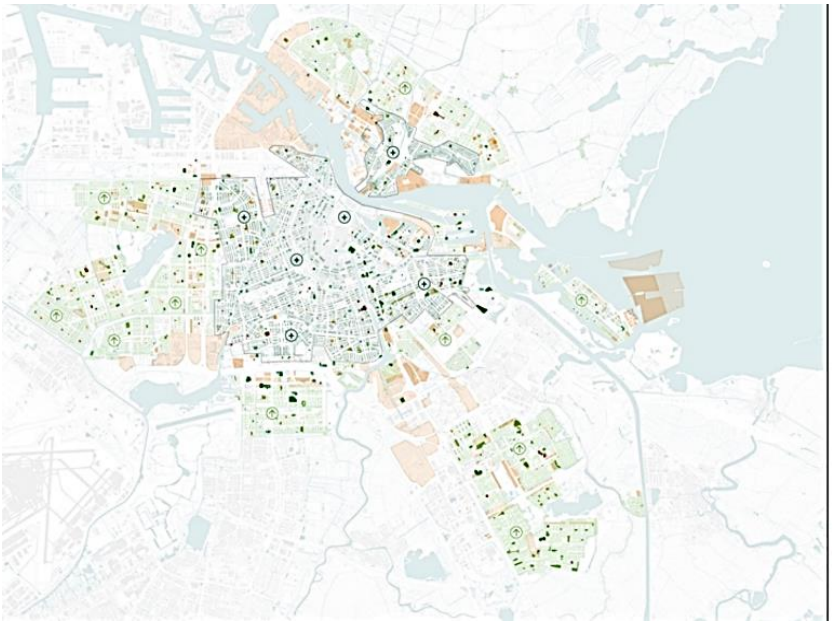


Figure 4-6: green neighbourhood scenario showing the projected 2050 infrastructure .

Figure 4-5: green metropolitan landscape scenario showing the projected 2050 infrastructure .

4.9.2 Comparing 2020 And 2050

The outcomes of the RQ4 analysis, comparing the 2020 PUGI-multiple regression results with the 2050-PUGI policy can for example showcase what policy has an expected significant effect on health-variables and what

policy-measures are not expected to improve health. This is relevant for bettering this 2050-PUGI policy, adding, or changing certain aspects in order to increase city-wide health benefits.

4.9.3 Expected Health impact of 2050 PUGI

The 2050 public urban green infrastructure in general is expected to improve health most where relatively most new green infrastructure is realized. In addition, the PUGI-policy that aims to include low-SES people is more beneficial than PUGI-policy that is aimed at already well-off people. The potential health-improvements are namely significantly higher for low-SES households and individuals. The spatial-level in which the PUGI is implemented is expected to be of additional importance, local PUGI and PUGI-connections are expected to improve health more compared to the addition of large-scale metropolitan PUGI. Local PUGI namely affect the low-SES people more as they tend to be less mobile, *figure 4-6* shows the 5 levels in which the 2050-PUGI is changing (Gemeente Amsterdam, 2020, p.24). Neighbourhoods that are currently already rich in PUGI but have low SES, are expected to be best benefitted from socio-economic improvements as, adding even more PUGI is showing diminishing returns.

Figure 4-7: five levels of new PUGI-implementation, from green buildings to urban-forests at the edge of the city

5. Analysis

5.1 Introduction

The analysis in many ways is a continuation of the theoretical framework and methodology. This is especially true for research questions 1, which is largely based upon the theoretical framework. The theory and findings from RQ1 are then used in RQ2 to find, filter, and pre-process the datasets that are relevant to answering the main research problem. In the subsequent RQ3, the pre-processed data is used to find statistical relationships between independent PUGI-variables, covariables, and the dependent health-variables. These regression outcomes are then interpreted using the statistical regression diagnostics. Finalizing this analysis is RQ4, where the statistical analysis from RQ3 will be compared with the spatial interventions in the PUGI that are expected to be implemented in 2050. *Table 5-1* gives an overview of the research questions, their goal, and the resulting chronology of the analysis steps.

Table 5-1: Research questions. analysis goals and relation between the RQ's

	RQ	*Goal*	*Input for next RQ*
1	*What theoretical relationships are found between PUGI-variables and health-variables?*	Review of existing related research to find what data, research methodology and findings are of importance.	Theory surrounding the relevant variables and relationships.
2	*What PUGI-variables, health-variables and co-variables are openly accessible (on a neighbourhood level in Amsterdam)?*	Obtain relevant open-data and pre-process this data.	Qualitative and relevant pre-processed datasets for the multiple spatial regression.
3	*What statistical relationships are found between PUGI and health variables among neighbourhoods in Amsterdam?*	Quantify the relation between the independent PUGI variables, the dependent health variables, and the covariables by doing a multiple spatial regression and interpret and validate the findings	Multiple regression results for showing how PUGI impacts different health variables, first part of answering the main research question.
4	*How can future public green infrastructure impact these relationships between PUGI-variables and health-variables?*	Create a 2050 PUGI-policy overview, analysing the potential impact given by the RQ3 (2020 situation) relationship on the 2050 situation variables.	Complete the answering of the main research question.

5.2 Research Question 1

5.2.1 Introduction

The goal of this research question is to analyse the existing theoretical framework and by doing so substantiate the relationships between PUGI variables, health variables and eventual covariables. The expected direction and strength of these relationships is required for creating a model that represents the real-world relationships properly. It is important no variables are overlooked, as this will make the multivariate regression model in RQ3 less fitting. In this literature review the expected relationships between

the independent and dependent variables are structurally presented to create a substantiation for the eventual model-variables, an overview is given in *figure 5-1.*

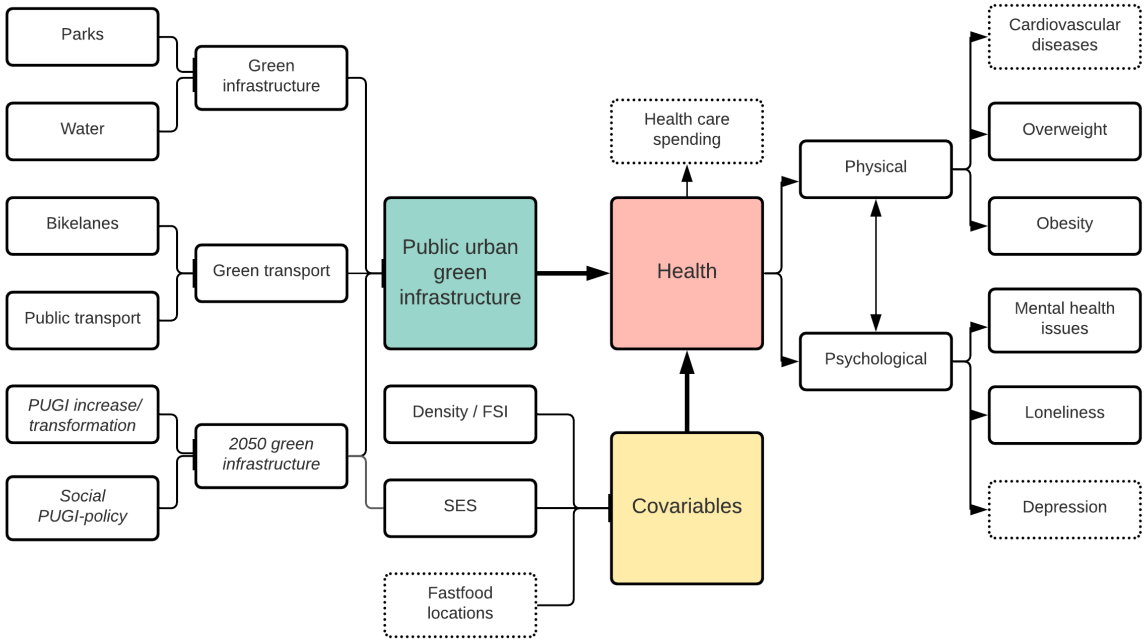


Figure 5-1: overview of interconnectedness of dependent and independent variables, dotted lines mean the variable is not available (aggregation discrepancy and/or non-open data)

5.2.2 Overweight, Obesity and BMI

Urban citizens in recent years on average experience a decline in physical activity and as a result an increase in more sedentary lifestyles (Congdon, 2019). This reduction in physical activity in combination with decreased offerings of healthy foods and increased offerings of unhealthy food, is often perceived as the main reasons obesity (BMI>30) and being overweight (BMI>25) prevalence is on the rise (Congdon, 2019). PUGI in theory could help decrease sedentary lifestyles and thus counter-act obesity to some degree. The reason for this is that PUGI is often used for physical activity and seduces people to use ac2tive modes of transport more often.

The municipality of Amsterdam, once every 5 years, looks into the types of activities that are undertaken in the municipal PUGI. The visitation-reason of a park was looked into by asking respondents what their main-activity was during their visit. The study found that the activities of park visitors in Amsterdam mainly consist of purely *physical activity* on the one hand (e.g. Walking (69%), cycling (17%) and running (24%)) and *relaxing* on the other hand (e.g. sitting in the sun (42%), enjoy the plants/animals (30%) and picnicking (27%)) (OIS, 2018). Where the first activities clearly directly impact sedentary lifestyles and thus has a direct impact on obesity, the latter one, however does this to a lesser degree. But *relaxing* most likely (however more indirectly) will also lead to less sedentary lifestyles as people move towards the park and often have more physical activity compared to *relaxing* at home. Both activities, physical/sporting and

social/relaxing furthermore increase social cohesion and reduce mental health issues, which in return lead to better physical health (Jennings & Bamkole, 2019). For this reason, the overweight *BMI>25* (%) and/or obesity *BMI>30* (%) variables are included in the research as the PUGI is expected to have a (in)direct impact on these variables.

5.2.3 Mental Health Problems and Loneliness

Besides PUGI having a relationship with important *physical* health variables like obesity and cardiovascular diseases it also has a relation with peoples mental health; mainly depression, stress and loneliness are affected by the presence of PUGI (Bos et al., 2020; Fan et al., 2011; Holtan et al., 2015; Jennings & Bamkole, 2019). As Holtan et al. puts it: '*Two European national scale studies found an independent positive relation between proximity to parks and (…) between green space and less loneliness and more social supports in the Netherlands'* (Holtan et al., 2015, p. 3). After this they state: *'(These findings are…) pointing to the importance of green space for populations that are less mobile and have less access to green space outside of their neighbourhood. These studies also demonstrate the need for a more fine-grained measurement of green spaces over a large geographic area.'* (Holtan et al., 2015, p. 3). Jennings adds to this on *page 3* of her article: '*People who are socially isolated tend to be less healthy and susceptible to stress, depression n, and cardiovascular issues'* (Jennings & Bamkole, 2019, p. 3). All of these authors are confirming the positive effect PUGI can have on mental health, and in addition this trickles down in order to improve physical health. In (Peen et al., 2010) they found that the more urban an area is the more psychiatric prevalence there is. And thus, directly implying a link between the availability of green and mental illnesses, where less green (urban areas) leads to more psychiatric health issues. For these reasons, the mental health variables *loneliness* (%) and *mental health problems* (%) are included in the analysis.

5.2.4 Healthcare Spending

As a result, urbanization and resulting decrease in PUGI also negatively impacts the efforts of the Amsterdam and Dutch government to reduce health care spending, a rapidly increasing budget-problem in many countries. With increased urban populations , urban related health problems will as a result grow in magnitude as they affect an increasing part of society, as more people live in these environment than ever before (Bos et al., 2020; Trencher & Karvonen, 2019). While the Dutch Gross Domestic Product (GDP) increased in the period 1998-2019 from €438 billion to €909 billion (+108%), health expenditure in the same period increased from €34 billion to €92 billion (+174%), corrected for population increase and inflation the net increase in health expenditure was +144% in this period, this is a discrepancy of 36 %-points with the GDP growth in 20 years (+144% health exp. vs. +108% GDP). With 92% of the Dutch people, and 99% of the Amsterdam people living in urban environments, an important part of the solution for this problem may lie in creating urban environments that facilitate better social, mental, and physical health. This research will try to find to what degree the PUGI affects the urban health (measured by several health variables), and if this relation is positive, solutions about health improvements may partly lie with increasing PUGI in urban areas.

This variable is not included in the research due to there being no available data on a neighbourhood level on health care spending, however the potential positive impact PUGI can have on reducing it makes the results more relevant as reducing health-care spending has society-wide benefits.

5.2.5 Covariables

Covariables, variables outside of the direct scope of the research that impact the relation between the initial independent and dependent variables, are important to consider. Urban density and socio-economic status are returning covariables in the existing research. These two covariables will be measured using well established metrics. The density example the floor space index and the per capita income. It is important to only measure 1 covariable using 1 variable, otherwise the covariables could mutually explain the same variance of the dependent variable, thus leading to multicollinearity. The covariables used in this research impact health variables due to them impacting *lifestyle* choices. Were for example a denser neighbourhood reduces car-usage and improve neighbourhood social cohesion and walkability. Higher SES in addition leads to better food intake, more budget for (outdoor) sporting activities, healthier jobs, and a better mobility (towards urban green). The combination of these lifestyle choices coming from the type of neighbourhood you live (dense/sprawl), and your SES (income/wealth) impact the health variables significantly and are therefore included as covariables.

5.2.6 Variable Overview

Table 5-2 (next page) gives an overview and theoretical substantiating of the required independent PUGI and dependent health variables that are preferably available as open data for the continuation of the analysis in RQ2. The PUGI, health and covariables that are used in this study are based upon a broad existing body of literature (Gemeente Amsterdam, 2017; Gunn et al., 2017; Nematchoua et al., 2020; Orstad et al., 2020; Paulin et al., 2019; Plane & Klodawsky, 2013; Staatsen et al., 2017; Vardoulakis et al., 2020; World Health Organisation, 2010, 2020a, 2020b). In *italics* are the variables that would have had added value, but due to unavailability on a neighbourhood aggregation level and/or non-openness of the data they are not used. The variables that are used in the multiple regression are in **bold.** Variables that are bold nor italic are variables that have the least theoretical substantiation and are merely included for possible usage later on.

Table 5-2: *Substantiated variables that are ideally used in this research, in red the variables that are not available due to aggregation discrepancy and/or openness criteria. Indep. = independent variable; Dep. = dependent variable.*

Independent variables (PUGI)	*Substantiation*
Dedicated bicycle paths (Indep. 1)	**(Plane & Klodawsky, 2013), (Vardoulakis et al., 2020), (Gemeente Amsterdam, 2017)**
Total public urban green (Indep. 2)	**(World Health Organisation, 2020a), (Staatsen et al., 2017), (Orstad et al., 2020), (Zanen et al., 2011)**
Total public urban blue infrastructure (Indep. 3)	**(Vardoulakis et al., 2020),** (Bos et al., 2020)**,** (Royal FloraHolland, 2021) **&** (Gemeente Amsterdam, 2020)
Surface area of parking spaces	(Paulin et al., 2019)
Trees	(Paulin et al., 2019), (Gunn et al., 2017). (Kumar et al., 2019) & (Mowafi et al., 2012)
Surface area of green roofs	(Paulin et al., 2019)

Public transport stops (Indep. 4)	(Jennings & Bamkole, 2019; World Health Organisation, 2010)
Fast food-locations	*(Mowafi et al., 2012), (Congdon, 2019) & (Lopez, 2004)*
Dependent Variables (health)	Substantiation
Overweight persons (BMI >25)	(Vardoulakis et al., 2020), (World Health Organisation, 2010)
Obese persons (BMI > 30) (Dep. 1)	(Vardoulakis et al., 2020), (World Health Organisation, 2010)
Loneliness (Dep. 2)	(Bos et al., 2020), (Holtan et al., 2015), (Jennings & Bamkole, 2019)
Mental health problems (Dep. 3)	(Bos et al., 2020), (Jennings & Bamkole, 2019)
Respiratory deceases	*(Paulin et al., 2019), (Vardoulakis et al., 2020), (World Health Organisation, 2010)*
Cardio-vascular diseases rate	*(Paulin et al., 2019), (Vardoulakis et al., 2020), (World Health Organisation, 2010)*
Average life expectancy	*(Staatsen et al., 2017), (World Health Organisation, 2010)*
Stress	*(Fan et al., 2011),* (Bos et al., 2020)*, (Holtan et al., 2015), (Jennings & Bamkole, 2019)*
Miscellaneous datasets	Substantiation
Current Neighbourhood population and population forecast (2050)	For averaging / general information
Neighbourhood surface area	For averaging / general information
Neighbourhood map (99 neighbourhoods)	(Nematchoua et al., 2020), (Staatsen et al., 2017)
Address density	For defining *urban* neighbourhoods
FSI-dataset (Covariable 1)	For defining *urban* neighbourhoods
CBS-vierkanten	Potential aggregation units
CBS socio-economic statistics (Covariable 2)	Use as covariable

5.3 Research Question 2

5.3.1 Introduction

The goal of RQ2 is to create a dataset for use as input for a multiple regression. The relevant data is joined to the neighbourhood-units for topology and policy-implementation purposes. In addition, the Neighbourhood is the natural aggregation unit many data are collected by. Meaning the data that can be aggregated to the neighbourhood-level without losing precision. Some variables are left out due to only being available in aggregation-units that have discrepancies with the neighbourhood-level that cannot be solved. Some health-variables for example are often not open and/or available on a neighbourhood level due to privacy reasons. The datasets need to be pre-processed to make them suitable for a multiple regression. This requires standardizing the data, (spatially) joining the data, combining the datasets, and doing a buffer analysis on the proximity of amenities such as green and blue infrastructure. In *figure 5-2 (next page)* the dataset construction is schematically presented and makes use of the in RQ1 substantiated variables. The dataset for tree-coverage is left out because it represents an incomplete overview of the actual tree-structure, something that is seen after overlaying the tree-dataset with a satellite map of the actual trees in Amsterdam.

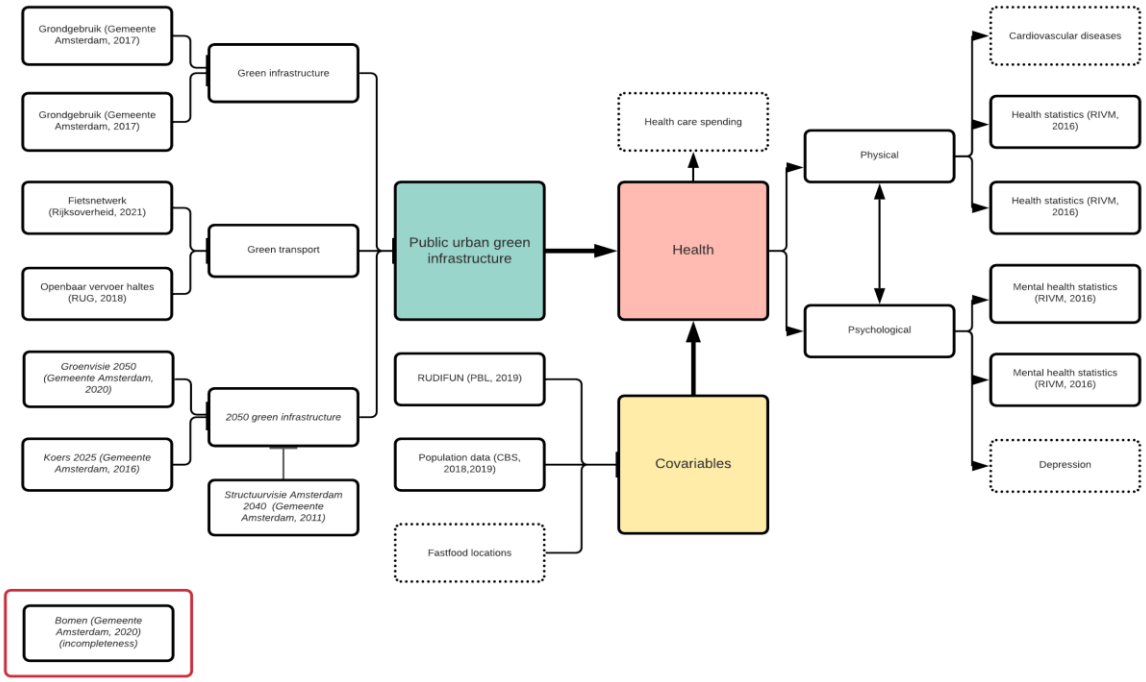


Figure 5-2: All datasets used for the variables, their year and source-holder/creator included. Dotted lines mean the data is not available/red outlines mean the data has shortcomings and therefore is not used.

5.3.2 Data Collection

Most of the variables mentioned in the RQ1 analysis are available via one of the 6 following (open) data portals*: 1. Nationaal Georegister, 2. CBS-statline, 3. data.amsterdam.nl, 4. RIVM-statline or 5. data.overheid.nl*. They are mostly available in the MIF file format and for that reason need to be changed to a shapefile (SHP) in order to make them suitable for working with them in ArcGisPro. This is done in FME-workbench. The resulting shapefiles are imported into ArcGisPro and inspected on errors or missing entries. The projections are kept to WGS84 Auxiliary Sphere (Zone 32N), as this is the projection that all municipal data is provided in, increasing interoperability between datasets. In FME-workbench it is ensured all datasets use this projection, and if not, are transformed to the WGS84 Auxiliary Sphere (Zone 32N) manually. This projection is suitable for the netherlands, ensures proper display of the maps in ArcGisPro and makes them interoperable with each other. All datasets that were found usable and were openly available for this study are shown in *table 5-3 (next page).* A extended version including metadata and pre-processing steps is available in *appendix 3.*

Table 5-3: The datasets that are obtained

Untranslated name (format) (type)	*Short description*
Functiemix (MIF) (Nominal)	Land use of the municipality of Amsterdam categorized by CBS_2 land use profiles (2017 data).
Fietsnetwerk Amsterdam (SHP) (Nominal)	Dataset showing all separated and marked bike-lanes in the municipality of Amsterdam (2016 data).
Groene daken (MIF) (Ratio)	All registered green roofs and, for most of them their surface area (2020 data).
Bomen (MIF) (Nominal)	Contains 4 datasets with all 259.431 public trees in Amsterdam and their location, type, and size. Private green infrastructure is not considered as they are not freely accessible and thus not part of the *public* green infrastructure. Think: Zoo, private gardens, private parking lot or industrial lot (2020 data).
Gebied Buurtcombinaties (MIF) (Nominal)	All 99 neighbourhoods (called *buurtcombinaties/wijken* in Dutch) (2020 data).
Bevolking wijken 2020-2050 (Excel) (Ratio)	The estimate of inhabitants per neighbourhood for the following years: 2020, 2025., 2030, 2040 and 2050. (2020 data)
Parkeervlakken (SHP) (Nominal)	Categorized database of all parking spaces in the municipality of Amsterdam; the categories differentiate between taxi parking, regular parking, handicap parking etcetera. This information can later be used to create the 2050 green infrastructure estimate (2017 data)
TOP10NL (SHP) (Nominal)	Detailed map of the ground features in the netherlands. For Amsterdam only the data from tiles 25W and 25O are needed.
Depression and loneliness statistics (Excel) (Ratio)	percentage of people that experience light depression and high depression symptoms + percentage of people with light loneliness and high loneliness symptoms (2016 data)
Obesity and overweight statistics (Excel) (Ratio)	Health variables and on a neighbourhood, district, and city level (2016 data)
Openbaar Vervoer haltes (SHP) (Nominal)	Point data of all public transport stops (bus, taxi, tram, ferry, and train) (2018 data)
RUDIFUN: Ruimtelijke dichtheden en Functie-menging Nederland (SHP) (Ratio)	Density data on different aggregation levels, from housing block to provincial. Data in net as well as gross FSI (2019 data)

5.3.3 Data Aggregation

To aggregate the number of spatial amenities with absolute locations, sizes, and shapes, such as PUGI, to each neighbourhood, while accounting for their proximity to the neighbourhoods, buffers are made. Buffers of 400 and 800 meter were chosen based upon literature. In (Nematchoua et al., 2020) 400 meter is seen as the ideal walking distance towards an amenity, including parks and canals, therefore this is seen as the minimal buffer distance to include. The municipality of Amsterdam aims to have a park within 10 minutes (800 meter) walking distance for every inhabitant (Gemeente Amsterdam, 2020), leading to the substantiation for the second buffer of 800 meter. Nørregaard et al. (2014) somewhat use the same buffers with respectifely 300 and 1000 meter, this implies the 400 and 800 meter buffers are fairly avergae in this type of research. '*Respondents living more than 1 kilometre away from green space were found to have higher odds of being obese compared with those living less than 300 metres from green space, after controlling for potential confounders'* (Nørregaard, Gram, Vigelsoe, Wiuff, & Birk, 2014, p.10). Finally (Congdon, 2019) uses a 0.5, 1 and 1.5 kilometre buffers were used, from which the 0.5 and 1 kilometre buffer are again an approximation of the 400 and 800 meter buffers used in this research, and (Fan et al., 2011) use a half-mile (800 meter) buffer as this is the area that is easily accessible by walking. *Figure 5-3* shows the scale of the buffer structure around the Neighbourhood Burgwallen Oude-Zijde (A00).

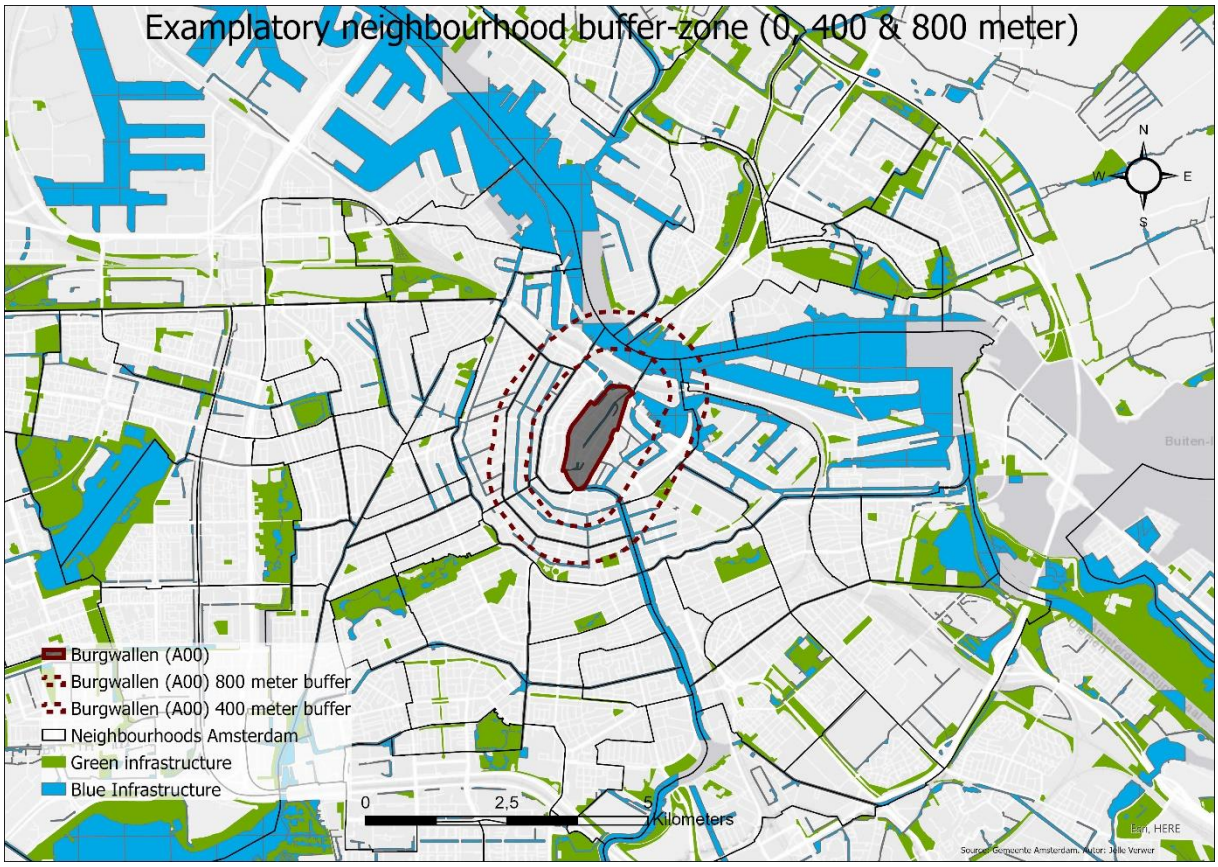


Figure 5-3: Exemplary buffer-structure of neighbourhood A00 (Burgwallen oude-zijde). The buffers are used to intersect the spatial features while weighting in a degree of proximity.

The buffers are *intersected* with all relevant spatial features to ensure all PUGI-features that are touching, overlapping and/or are within the service area of the buffer are included in the neighbourhood PUGI-dataset. This is done due to the PUGI being at least partly within the neighbourhood proximity can be expected to be used as a whole. Imagine the entrance of a park being 10 minutes from your home, this will mean the amenities of the whole park fall within the 10 minute buffer as from that point on the park is at your disposal completely.

5.3.4 Data Standardization

The data is combined in a spatial dataset, besides the green, blue and *green-t*ransport (public and cycle) infrastructure also the FSI (density), low income individuals (%), and several (mental) health variables are included. The complete table with all data that is used is shown in a separate file due to the size of the table (X34*Y91), *(Thesis dataset Jelle Verwer)*. The data that has some relation with size (of a neighbourhood in this case) is first averaged out and then standardized using the formula $z = \frac{x_i - \bar{x}}{s}$, where z = standardized score, and x = the original value, $\bar{x}$ = average of the variable and s = standard deviation of the variable. This ensures the outliers are better comparable. All data whose values are influenced by the neighbourhood size are standardized. Looking into the distribution of a spatial phenomenon (e.g., Surface area of parks & number of trees etc.) will lead to the larger neighbourhoods being in general more saturated with PUGI, thus these values are also divided by the neighbourhood size before standardizing. .

5.3.5 Health Variables

The health data that is ideally used for the health variables sadly not all is open data. Some important variables are openly found on the geoportal of the *Rijksinstituut voor Volksgezondheid en Milieu* (RIVM), the official Dutch governmental institute for Public Health and the Environment, these open datasets include BMI, depression, and anxiety on a neighbourhood level. But other health-related datasets are not available (on a neighbourhood level), for example cardiovascular and respiratory disease rates. This limitation leads to these health variables not being used. On the RIVM data portal (*RIVM Statline)* health data on neighbourhood level is available for some variables. The neighbourhood IJburg-Oost (M50) has no available data from RIVM statline as the only Amsterdam neighbourhood as there are no inhabitants still, therefore this neighbourhood is not considered in the research. But this was not done anyway, namely also due to it being not urban according to the CBS-classification (based upon address-density) in combination low with FSI-values as mentioned previously.

5.3.6 Covariables

Covariables are important variables that have significant impact on the spatial correlation between PUGI-variables and health-variables *(table 5-4)*, but that are outside of the direct scope (thus no part of the PUGI-variables). First of all the socio-economic factors within each neighbourhood have expected correlation with health-variables, this relation is studied and substantiated by (Plane & Klodawsky, 2013) and (Mouratidis, 2020). Some variables that in previous studies impact health variables on a neighbourhood level are floor space index (Density), address density (Density) car possession (SES), fast-food locations, income (SES), property value (SES), and social-security recipients (SES). It is for example found that low-income groups in the Netherlands have a 17% obesity rate (BMI > 30) while high-income groups only have a 9% obesity rate (Staatsen et al., 2017). This implies that low income areas in advance are disadvantaged in this dependent variable, independent from the PUGI.

Table 5-4: covariables that are found to be correlating with health-variables

Dataset name (format) (type)	*Source (weblink)*	*Contents & info*
Address-density (Excel) (Ratio)	https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84583NED/table?ts=1605621440710	- Address-density per square kilometre for all 99 neighbourhoods in Amsterdam (2019 data).
Socio-economic data (Excel) (Ratio)	https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83765NED/table?ts=1605626906369	- Average property value; per (working) person income; government assistance (total persons) (2017 data).
Floor Space Index (FSI) and Ground Space Index (GSI) (Ratio)	http://data.overheid.nl/dataset/1b104806-a4dc-4438-9559-7149c1b93d10	- FSI per street, neighbourhood, municipality (2017-2019 data).

The Floor space index is a useful measure for defining the degree of urban density and on that basis include and exclude certain neighbourhoods. Within the Dutch FSI dataset, made available by *het Planbureau voor de Leefomgeving (PBL)*, their seems to be many faulty entries. FSI scores of up to 10.000 are present in the dataset. This seems to be an error coming from the fact that the BAG, the database from which the building footprint is extracted has information on the footprint of buildings that is wrongly interpretated by the PBL. Metro-stops for example sometimes have a footprints as small as 3m². This leads to the building's actual floor space (say 3.000m²) being divided by a footprint of 3m², leading to a 1000 FSI-score. For this reason, the highest known reasonable FSI (Singapore, 25), is used as a maximum for the dataset. Using the FSI on a block level, a map is made, showing the FSI in great detail (figure 5-5). Higher

FSI-densities are observed in the central 17th/18th/19th/early 20th century neighbourhoods, as well as in the more recent southern IJ-waterfront and IJburg-area.

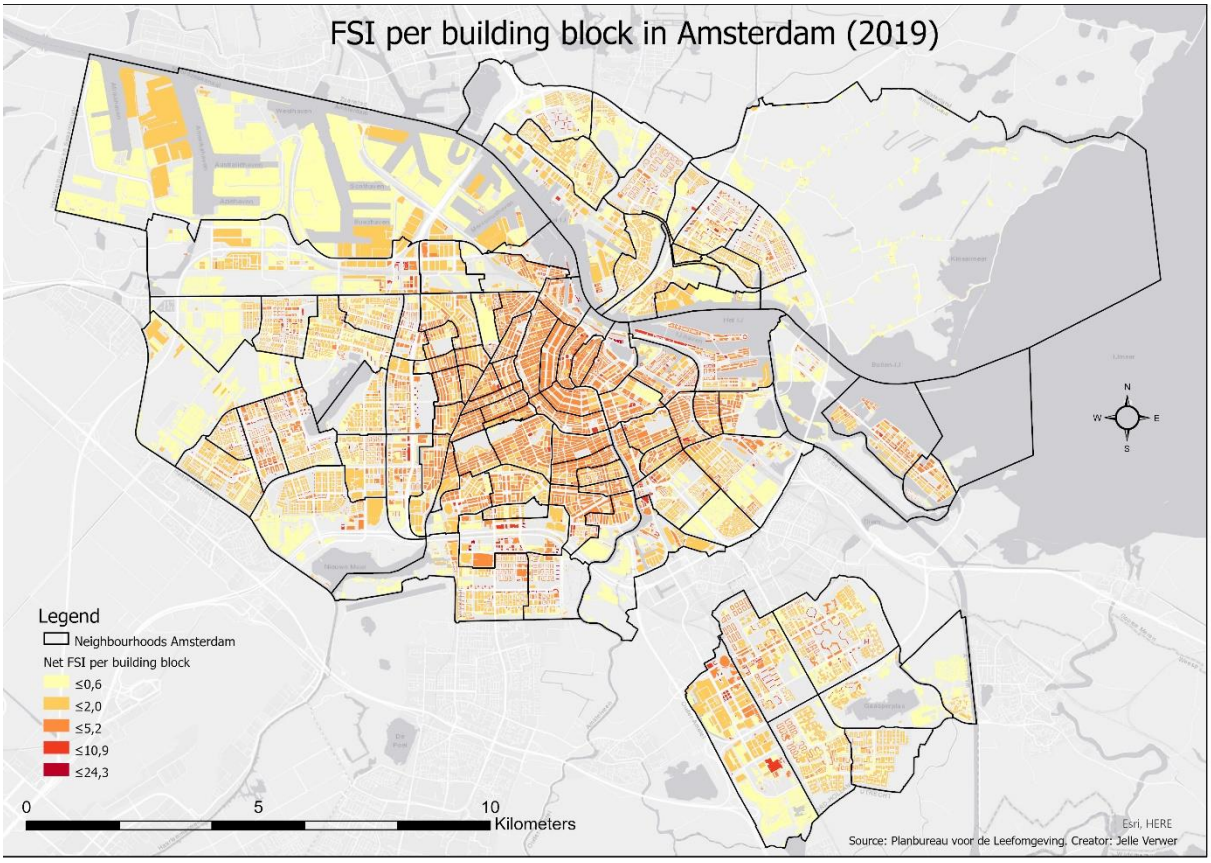


Figure 5-5: the FSI on a housing-block level in Amsterdam (PBL, 2019)

In addition to the FSI, the address density is used to rank the urbanity of neighbourhoods. The address density should, according to the CBS, be above 1.500 addresses per km² (CBS, 2020b) to be seen as *mildly urban,* and above 2.500 addresses per km² to be considered *highly urban*. For the purpose of this research a minimal address density of 1.500 per km² will be used. Narrowing the scope to public *urban* green infrastructure. Taking into consideration the 2020 official demographic statistics from CBS statline (CBS, 2020a), the neighbourhood address density is <1.500 in 8 of the 99 neighbourhood, between 1.500 and 2.500 in 10 of the 99 neighbourhood and above 2.500 in the remaining 81 neighbourhoods. On average the address-density in Amsterdam is 5.613 with a stretch between 368 and 12.218. It is after some inspection of the dataset clear that the 8 neighbourhoods with a low address density (<1.500) in general have a specific function, development status or location that makes them not usable for the urban scope of the research, an overview is shown in *table 5-5* and *figure 5-6 (both next page).*

Table 5-5: description of the 8 of non-urban neighbourhoods according to CBS address-density classification (2020)

Neighbourhood code	*Neighbourhood name*	*Address density (2020)*	*Reason for low address-density*
B10	Westelijk Havengebied	715	Industrial area/port area
F11	Bedrijventerrein Sloterdijk	1.435	Transport hub (public transport) + many (large) hotels and office buildings.
F80	Lutkemeer/ Ookmeer	1.100	Western edge of city where rural meets urban, a lot of farmland and low-density residential area
M34	Zeeburgereiland /Nieuwe Diep	1.119	Residential area in development (inhabitants: 4.704 (2020) → 17.193 (2050))
M50	IJburg-Oost	1.123	Residential area at start of development (inhabitants: 0 (2020) → 21.861 (2050).
T92	Amstel III/Bullewijk	1.469	Residential area in development (inhabitants: 742 (2020) → 22.500 (2050))
T98	Driemond	452	South/eastern edge of city where rural meets urban, a lot of farmland and low-density residential area
N73	Waterland	368	Northern edge of city where rural meets urban, a lot of farmland and low-density residential area

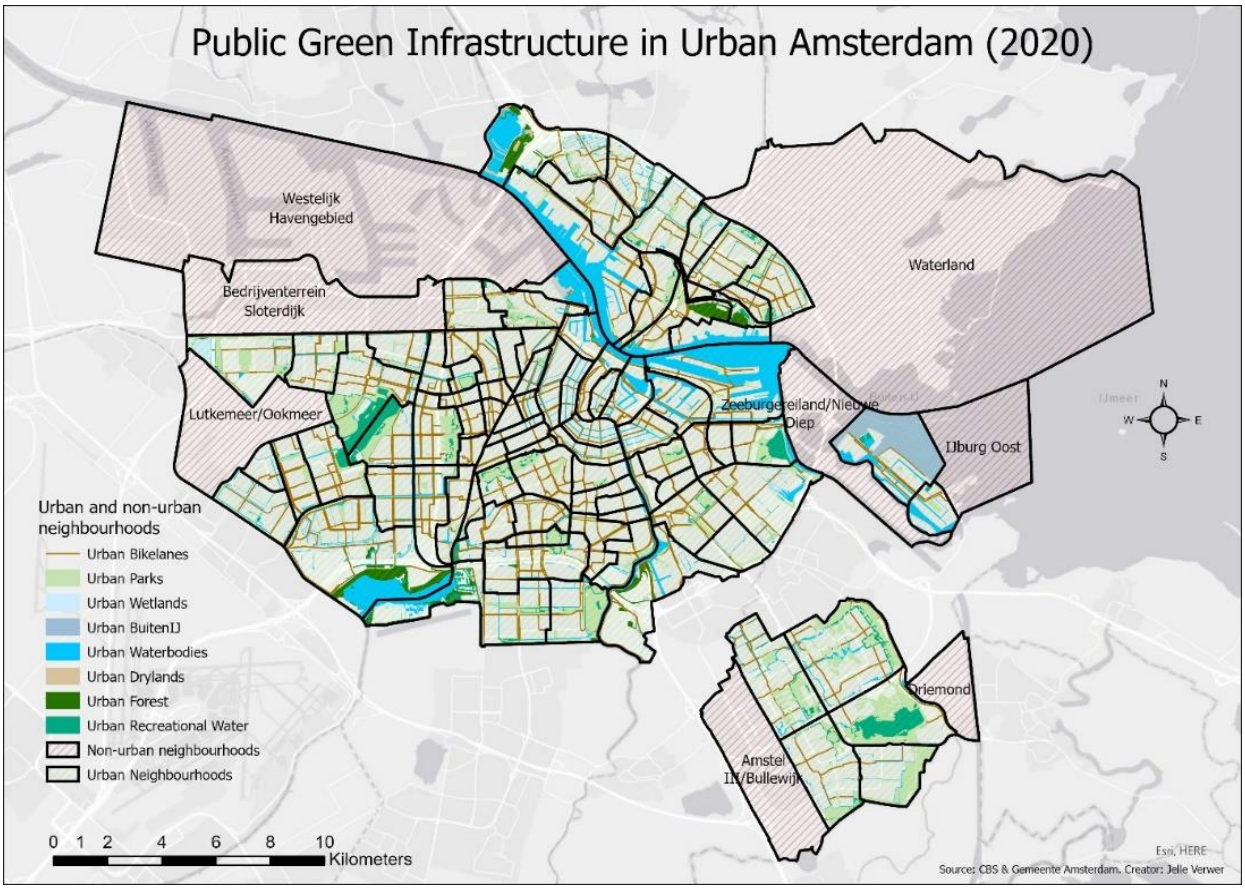


Figure 5-6: Public urban green infrastructure in Urban classified neighbourhoods

5.4 Research Question 3

5.4.1 Introduction

The multiple-regression analysis uses the workflow seen in *figure 5-7*, from which the green squares are corresponding with RQ2 to show the connectiveness with RQ3 better, and the yellow squares correspond to RQ4. The analysis uses a multiple regression model. The model-hypothesis is that an Ordinary Least Squares regression (without spatial weights) does not account for the heteroscedasticity in most of the variables and therefore a regression model with a spatial lag/error-coefficient is better suitable. All three of these regression models generate 1 coefficient for all neighbourhoods, but the lag and error-model in addition account for the interdependency between neighbours (the neighbourhood itself can be based upon topology, proximity of other factors). The analysis null-hypothesis that will be tested is that the green and blue infrastructure in combination with the covariables (FSI and Income) are *no* significant predictors to obesity (and in a later stage also mental health problems). This null-hypothesis can be rejected when the slope of the regression curve is below or above 0 (and thus there is some sort of positive or negative relation between the explanatory variables and the dependent variable) *and* the p-values of the variables are significant, having $p < .05$, meaning there is a <5% chance the null-hypothesis is correct, and the variable does not correlate with the dependent variable. The threshold-probability of .05 is a widely accepted balance between type-1 and type-2 errors, and thus for accepting the null or alternative-hypothesis incorrectly.

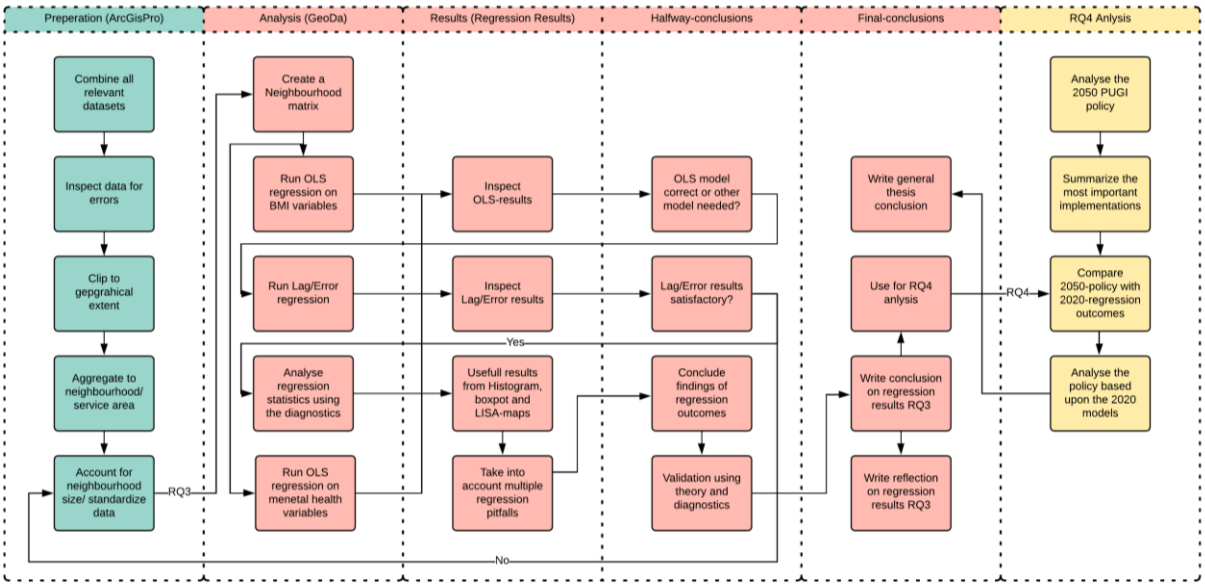


Figure 5-7: Workflow for analysis RQ3 (including the final RQ2 and first RQ4 steps for completeness)

The analysis of RQ3 will consist of 9 distinguishable multiple regression-analyses to cover the 3 dependent variables in relation to the 3 aggregation levels. The variables are % of people with a BMI>30 (BMI30), % people with medium mental health problems (MMHP%) and % of people who consider themselves as lonely (LO%). These 3 health variables will be modelled in relation to PUGI aggregated to 3 neighbourhood classifications: Neighbourhood, Neighbourhood+400 and Neighbourhood+800. All models

will include 2 important covariables; density (FSI) and % people with a low income (LINC%). The order of the regression will be top to bottom and left to right (1A, 2A, 3A, 1B, 2B, 3B, 1C, 2C, 3C). The analysis from tulip 2 and 3 (buffered units) will be explained more briefly except for when they add significantly to the relative model-fit, model-R^2 and/or probability of the model-coefficients. All 9 model-combinations are shown in *table 5-6.* An overview of the variables and their abbreviation-translation table is available in *appendix 6.*

Table 5-6: overview of the 9 regression analyses, OLS-models will use a K5 matrix while lag/error-models will use a Q1 matrix

	Aggregation / *Dep. Variable*	*1*	*2*	*3*
		Neighbourhood	Neighbourhood + 400 meter buffer	Neighbourhood +800 meter buffer
A	BMI>30 (%) *(BMI30)*	GRNEIST BLNEIST BINEIST PTNEIST LINC% FSI	GR400ST BL400ST BI400ST PT400ST LINC% FSI	GR800ST BL800ST BI800ST PT800ST LINC% FSI
B	Medium mental health issues (%) *(MMHP)*	GRNEIST BLNEIST BINEIST PTNEIST LINC% FSI	GR400ST BL400ST BI400ST PT400ST LINC% FSI	GR800ST BL800ST BI800ST PT800ST LINC% FSI
C	Loneliness (%) *(LO%)*	GRNEIST BLNEIST BINEIST PTNEIST LINC% FSI	GR400ST BL400ST BI400ST PT400ST LINC% FSI	GR800ST BL800ST BI800ST PT800ST LINC% FSI

5.4.2 Variable Exploration

For clarification and transparency purposes an overview of all used variables is given in *appendix 5*. Using these pre-processed variables, the first step in the multiple regression analysis is inspecting the data and see how the dependent variable is distributed in space (potential spatial autocorrelation); this can be done using a global Moran's' I scatterplot as seen for BMI in figures 5-8. This scatterplot uses a K5 neighbourhood matrix, the X-axis shows the distribution of the residuals and the *lagged* Y-axis the residuals value of the K5-neighbours. Thus, showing if values are surrounded by other similar values (SA). Clearly seen is the clustering of high-high (hotspots) values and low-low (coldspots) values with BMI>30 (*figure 5-8).*

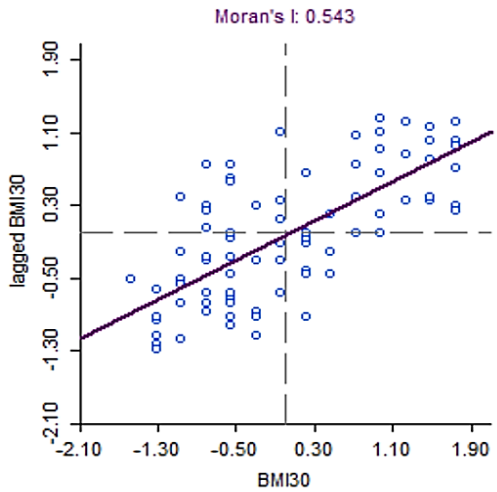


Figure 5-8: strong positive spatial autocorrelation of BMI>30 prevalence (K5)

Even though, first inspection of the dependent variable, suggest a multivariate regression using lag

and/or error-coefficients is expected to be most fitting, a standard non-spatial OLS regression is firstly ran as this is the simplest type of regression, and when it has a great fit with the data, a spatial matrix is cumbersome and merely adding complexity to the model. *'In general, but not a rule, measures of strong model fit (e.g. R^2 > 0.8), coupled with weak and insignificant levels of spatial autocorrelation in the residuals, suggest that a linear regression would be appropriate'* (Comber et al., 2020, p.10). Thus, when the OLS model-fit is high, and the SA in the residuals is insignificant, the OLS-model type is most fitting.

To account for SA, a weight matrix is incorporated to make it possible to adjust the model based upon the assumption that the coefficient should include this interdependency (by means of a lag or error-coefficient). The neighbourhood-structure is analysed first in ArcGisPro, showing that on average a neighbourhood has ±5.2 neighbours with a standard-deviation (σ) of ±2.0. This analysis makes use of the topology of the neighbourhoods and thus tells us on average slightly over 5 neighbourhoods are touching the border of every neighbourhood. The distribution of neighbours has a close to normal distribution with a

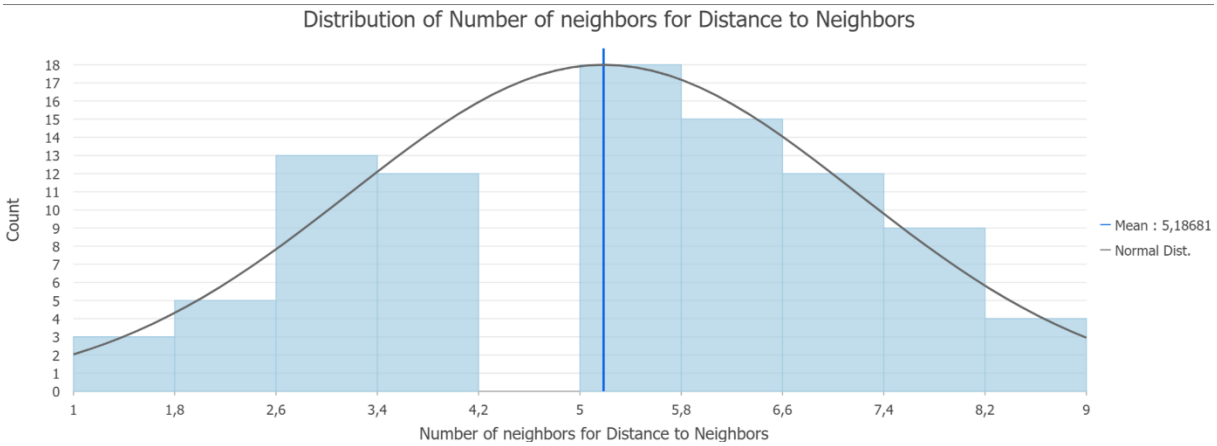


Figure 5-3: Distribution of neighbours based upon queens-case topology (touching edges & corners).

large part of the samples falling between the 3-7 range (average ± σ), however the sample-size (91) is not large enough for a smooth curve. See *figure 5-10* for visualisation.

5.4.3 OLS-model: Obesity

First, a OLS-model without a matrix is carried out. Including four explanatory PUGI-variables and 2 covariables, namely: green infrastructure, (*GRNEIST*), blue infrastructure, (*BLNEIST*), bike infrastructure (*BINEIST*), public transport infrastructure (*PTNEIST*) low-income people (%) (*LINC%*) and floor space index (*FSI*). The dependent variable is obese people (*BMI30*), the result is seen in *table 5-7 (Next page; Appendix 9)*. The *adjusted* R^2 is *.663955*, meaning 66.4% of variation in obesity (dependent variable) is in theory determined using the explanatory variables in this OLS model.

Table 5-7: Model-diagnostics, OLS test without spatial weights (BMI>30) *(Appendix 9)*

Test	*Value*	*Probability (ρ)*
Adjusted R²	.663955	*N.A.*
F-statistic	30.6369	3.22438e-19
Log likelihood	-200.442	*N.A.*
Akaike info criterion	414.884	*N.A.*
Schwarz criterion	432.46	*N.A.*
Multicollinearity condition number	8.589815	*N.A.*
Breusch-Pagan	27.0301	.00014
Koenker-Bassett	19.8061	.00300

Variable	*Coefficient (β)*	*Std. Error (σ)*	*T-statistic*	*Probability (ρ)*
CONSTANT	10.116	.980984	10.3121	.00000
BINEIST	.137277	.316155	.434208	.66525
BLNEIST	.453253	.296908	1.52658	.13062
PTNEIST	-.724666	.290463	-2.49487	.01456
GRNEIST	-.429561	.269973	-1.59113	.11534
LINC%	.501862	.0564362	8.89256	.00000
FSI	-4.59325	.664701	-6.91025	.00000

Inspecting the regression-coefficients, it is seen that green infrastructure is negatively insignificantly correlated to obesity (*β -.429561; ρ .11534)*. For Blue infrastructure, the coefficient is positive but also insignificant (*β .453253; ρ .13062)*. When looking at the public transport stops (*β -.724666; ρ .01456)*, they seem to be having a negative, but significant relation with obesity, in theory meaning more bus& tram-stops lead to less obesity. Bike infrastructure seems to be a highly insignificant predictor of obesity *(ρ - .66525),* with a slightly positive relation (*β .137277).* The covariables, low-income people and FSI, are both highly significant (*ρ .00000).* Whereas low-income people tend to increase the percentage of obesity (*β .501862*), a higher FSI (density) tends to lower the percentage of obesity *(β -4.59325)*. The coefficients in this case also depict the BMI, meaning in *theory* one could *calculate the expected* BMI (meaning to *predict* 66.4% of the BMI-variance) in a neighbourhood.

Looking at the relative model fit, by means of the *Log-likelihood, Akaike info criterion (AIC)* and *Schwarz criterion (SC)* indicators, allows for future relative model-comparison (these indicators allow to compare the balance between model-simplicity and explanatory power). Models with a low AIC and SC, and a high Log-likelihood are favourable. These indicators are relative, and thus the value itself has no meaning unless another value of these indicators is compared to them.

The multicollinearity condition number (MCN) indicates if two or more variables (dependent and/or independent) have too strong of a correlation in their explanatory factor. The MCN should be well below 30, which is the case in this model (*8.589815)*. Meaning there are no explanatory variables and/or covariables that have too much correlating effects regarding the explanation of the dependent variable.

Heteroskedasticity is measured using the Breusch-Pagan (BP) test and/or Koenker-Basset (KB) test. These tests show if the residuals of the independent and/or dependent variables are constant over all samples. If these tests are significant the explanatory variables and their relationships with the dependent variable are non-stationary. This in return means the explanatory variables may explain large parts of the

dependent variable in one sample, while barely explaining the relation in other samples. Heteroskedasticity results in the standard-error of the coefficients being over or under-estimated. Both indicators have a probability well below *.05 (≈ 0),* meaning there is in fact significant heteroskedasticity.

5.4.4 OLS-model: Obesity (K5)

Spatial weights (K5) are used in this regression to inspect the SA and, if needed use a different model-type (Lag/Error). In addition, the highly insignificant bike-infrastructure *BINEIST* (ρ *.66525*) is excluded. The resulting regression report is shown in *table 5-8 (Appendix 10).*

Table 5-8: Model-diagnostics, OLS test with K5 spatial weights and without BINEIST (BMI>30) *(appendix 10)*

Test	*Value*	*Probability (ρ)*
Moran's I	4.9438	.00000
Lagrange Multiplier lag (Robust)	27.2619 (10.6738)	.00000 (.00109)
Lagrange Multiplier error (Robust)	16.9584 (.3703)	.00004 (.54284)

Test	*Value*	*Probability (ρ)*
Adjusted R²	.667163	*N.A.*
F-statistic	37.080	5.47822e-020
Log likelihood	-200.544	*N.A.*
Akaike info criterion	413.088	*N.A.*
Schwarz criterion	428.153	*N.A.*
Multicollinearity condition number	8.414575	*N.A.*
Breusch-Pagan	22.6123	.00040
Koenker-Bassett	16.2072	.00628

Variable	*Coefficient (β)*	*Std. Error (σ)*	*T-statistic*	*Probability (ρ)*
CONSTANT	10.1945	.959548	10.6243	.00000
BLNEIST	.428214	.289861	1.47731	.14329
PTNEIST	-.654161	.239686	-2.72924	.00771
GRNEIST	-.379809	.243282	-1.56119	.12220
LINC%	.497653	.0553312	8.99407	.00000
FSI	-4.61872	.658939	-6.91025	.00000

This model is similar to the first non-weighted OLS, even though this time the bike-infrastructure is excluded as an explanatory variable. Regarding the coefficients and probabilities only small changes are made to the model. The heteroskedasticity indicators stay similarly significant but are lower on average, the relative model-fit is better overall. The adjusted R-squared is slightly higher and the multicollinearity is lower. Overall, this model has a better fit and better diagnostics. Most of the diagnostics for SA are significant, except for the robust Lagrange multiplier (lag) test; meaning a OLS may not be the best fit, and a lag-model should be considered. A lag-model calculates an extra coefficient that considers the neighbouring values (using a non-adaptive matrix like a queen's contiguity). The lag-coefficient, by doing so, muffles the clusters to lower the overall residuals.

5.4.5 Spatial-Lag-Model: Obesity (Q1)

The lag model adds to the OLS-model a lag-coefficient (*W_BMI30)* that uses the Q1 contiguity matrix to calculate a coefficient to weight the neighbours with. By doing so the extreme values (outliers) get levelled out by their neighbours and extreme values themselves increase the value of their neighbours to create, in

theory, a smoother regression line with less errors. The resulting diagnostics are seen in *table 6-8 (Appendix 11).*

Table 5-9: Model-diagnostics, lag-model with Q1 spatial weights and without BINEIST (BMI>30) *(appendix 11)*

Test	Value	Probability (ρ)
R^2	.789065	N.A.
Log likelihood	-185.158	N.A.
Akaike info criterion	384.316	N.A.
Schwarz criterion	401.892	N.A.
Breusch-Pagan	24.6292	.00016

Variable	Coefficient (β)	Std. Error (σ)	z-value	Probability (ρ)
W_BMI30	.486869	.0788492	6.17469	.00000
CONSTANT	3.38542	1.27019	2.66527	.00769
BLNEIST	.172859	.235551	.73385	.46304
PTNEIST	-.609779	.190844	-3.19518	.00140
GRNEIST	-.252123	.193128	-1.30547	.19173
LINC%	.433223	.046151	-4.70025	.00000
FSI	-2.74551	.58412	-4.70025	.00000

The *W_BMI30* in this table means that the expected BMI for a certain neighbourhood is increased by *.486869 ** the weighted BMI of its neighbours (*βw*). This βw is adding to the regular model coefficients who are now overall lower due to this lag-BMI coefficient. The *βw* is not a constant factor as some neighbourhoods with this queen's contiguity have more neighbours than others. The lag-model has a better fit according to the relative model-fit indicators (AIC, Schwarz & log-likelihood). The R-squared increased to *.789065* (compared to R^2 *.667163* in the OLS). The probability of the explanatory variables (based upon the z-value instead of the t-statistic) is still insignificant for blue and green infrastructure. But significant for the *W_BMI30*, *CONSTANT*, *PTNEIST*, *LINC%* and *FSI*. Meaning these explanatory variables, with inclusion of a lag-coefficient, are more significant predictors for obesity. Concludingly, the green and blue infrastructure, ate least on a neighbourhood-level does not seem to be a strong and/or significant predictor of obesity. The covariables (and public transport stops) however are significantly correlated to obesity prevalence.

5.4.6 Buffered Models: Obesity

Expanding the aggregation of the PUGI to a *Neighbourhood+400m* buffer *(Appendix 12)* leads to less satisfactory results. The buffered OLS-model results in a lower adjusted R^2 (*.626090*) and a lower model fit within all indicators (*AIC 424.6; log-likelihood -205.4; SC 442.176*). In addition, the PUGI-variables are all highly insignificant (*GR400ST ρ.48775; BL400ST ρ.39023; PT400ST ρ.99204; BI400ST ρ.68510*), while the covariables stay highly significant (*LINC% ρ.00000; FSI ρ .00000*). The SA-statistics indicate a spatial lag-model is most suitable.

The *Neighbourhood+400m* spatial lag-model *(Appendix 13)* has a higher R^2 (*.762863*), compared to the OLS-model of the same aggregation (*.626090*), but still a lower R^2 *compared* to the *Neighbourhood* spatial lag-model (*.789065*). However, the coefficients are highly insignificant (*GR400ST ρ.98187; BL400ST ρ. 98659; PT400ST ρ. 68226; BI400ST ρ. 93695*), meaning the chance the null-hypothesis is true (the 400-meter PUGI

has no relation to obesity) is highly significant. The covariables and lag-coefficient are highly significant again with (*LINC% ρ.00000; FSI ρ.00000*). The lag-BMI30 (*βw*) is also highly significant with *ρ.00000*.

Expanding the aggregation of the PUGI to *Neighbourhood+800m* buffer results in the following model-diagnostics *(Appendix 14).* The OLS-model results in a lower adjusted R^2 (*.628083*) and a lower model fit within all relative indicators (*AIC 424.113; log-likelihood -205.057; SC 441.69*). In addition, the PUGI-variables are, again, all insignificant (*GR800ST ρ.13499; BL800ST ρ.73915; PT800ST ρ.37567; BI800ST ρ.35783*), while the covariables stay highly significant (*LINC% ρ.00000; FSI ρ .00000*). The SA-statistics also indicate a spatial lag-model would be most fitting.

The *Neighbourhood+800m* spatial lag-model (*Appendix 15*) has a higher R^2 (*.763369*), compared to the OLS-model of the same aggregation (*.628083*), but still a lower R^2 compared to the *Neighbourhood* spatial lag-model (*.789065*). However, the coefficients are highly insignificant again (*GR800ST ρ.35646; BL800ST ρ. 83808; PT800ST ρ. 48481; BI800ST ρ. 51740*), meaning the chance the null-hypothesis is true (the 800 meter PUGI has no relation to obesity) is highly significant. The statistics show however, small improvements compared to the 400-meter buffer model. The covariables are highly significant with (*LINC% ρ.00000; FSI ρ.00000*). The lag-BMI30 (*βw*) is also highly significant with *ρ.00000*.

Concludingly a spatial lag-model is within all 3 aggregation areas the best model-choice based upon the OLS SA-statistics. The larger aggregation units (+400 and +800 buffer) add little statistical significance, explanatory power, or model-fit, compared to the non-buffered regression-model. The models that were produced with the buffers had lower model-fit statistics and low-significance in all PUGI coefficients. Meaning there is a realistic chance the null-hypothesis (the PUGI has no correlation with the health variables) is true for the 400 and 800 meter buffers. These aggregation units have little significant statistical relation with the obesity rates in neighbourhoods. The over-arching trend in all 3 models is the importance of the covariables FSI and low-income persons. These variables both have significant impact on the expected obesity prevalence.

5.4.7 OLS-Model: Mental Health Problems (K5)

To quantify the relation between PUGI and the percentage of people with mental health problems on a neighbourhood level (MMHP%), the PUGI-variables, in combination with the 2 covariables (FSI and low-income%) are modelled. Using an OLS-model, and with inclusion of K5 spatial weights the OLS-regression model SA-indicators can be used to decide if another model type might be more fitting. The regression indicators and coefficients of this model are shown in *table 5-9 (Next page, Appendix 16) .*

Table 5-10: Model-diagnostics, OLS test with K5 spatial weights (Mental health problems) *(Appendix 16)*

Test	*Value*	*Probability (ρ)*
Moran's I	.8243	.40979
Lagrange Multiplier lag (Robust)	1.8117 (2.0285)	.17831 (.15438)
Lagrange Multiplier error (Robust)	.1107 (.3274)	.73938 (.56717)

Test	*Value*	*Probability (ρ)*
Adjusted R²	.735102	*N.A.*
F-statistic	42.6256	1.76471e-023
Log likelihood	-209.693	*N.A.*
Akaike info criterion	433.387	*N.A.*
Schwarz criterion	540.963	*N.A.*
Multicollinearity condition number	8.589815	*N.A.*
Breusch-Pagan	20.9257	.00189
Koenker-Bassett	6.3373	.38649

Variable	*Coefficient (β)*	*Std. Error (σ)*	*T-statistic*	*Probability (ρ)*
CONSTANT	39.747	1.08596	36.6007	.00000
BLNEIST	-.837377	.349987	-.903694	.36874
PTNEIST	-.00837377	.321546	-.0264749	.97894
BINEIST	-.316281	.349987	-.903694	.36874
GRNEIST	.44666	.298864	1.49453	.13879
LINC%	.943945	.0624756	15.109	.00000
FSI	-3.2319	.735832	-4.39217	.00003

This OLS model has a strong absolute fit, with a adjusted R^2 of *.735102*. The coefficients and associated significance show that the *GRNEIST*, *BINEIST* and *PTNEIST* variables are not significantly correlating with mental health problems, *BLNEIST* is however negatively significantly correlated; more neighbourhood blue infrastructure in theory lowers the prevalence of mental health problems. The probability *GRNEIST*, *BINEIST* and *PTNEIST* are not significant in relation to mental health problems (and thus that the null-hypothesis is true) is ranging from *.13879 (GRNEIST)* and *.36874 (BINEIST)* all the way to *.97894 (PTNEIST).* The covariables are highly significantly correlating to the dependent mental health variable (*FSI .00003; LINC% .00000*). Whereas more low-income people tend to higher the prevalence of mental health problems, a higher FSI tends to lower it. The constant is in addition highly significant (*p .00000*).

Multicollinearity is not strong and thus the variables do not show problematical correlation in their explanatory contribution. Looking into the heteroskedasticity of the residuals; the BP-test is significant, while the KB-test is not. The visualisation of the heteroskedasticity show the residuals are very dispersed in some neighbourhoods, but not per se skewed to one side of the x-axis. This makes it disputable if there is actually heteroskedasticity (*figure 5-11)*. None of the diagnostics for SA are significant, for example the global Moran's I has a probability that lies well above the *.05* significance threshold (*.44121*). These spatial dependence diagnostics in addition show the OLS-model has in theory the

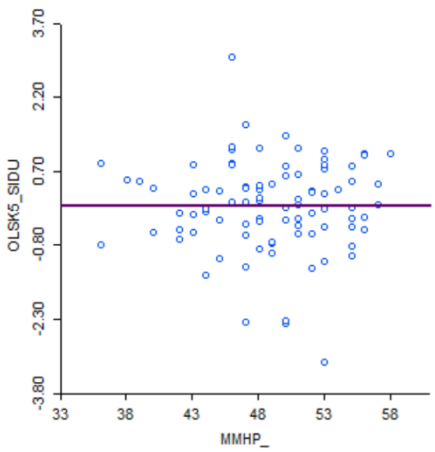


Figure 5-11: Potential heteroskedasticity in the mental health problem residuals

highest fit in modelling the spatial relation between PUGI and health problems. Using a spatial-lag or spatial-error model have no expected impact on the (relative) model fit. The high model-fit (*Adjusted R^2 .735102*) confirms that the OLS-model has good fit to the data.

A new OLS-model is made excluding the highly insignificant variables *BINEIST* and *PTNEIST* (*appendix 17).* These explanatory variables are excluded from the model to inspect if there are any improvements in the other diagnostics by doing so. The model fit is slightly lowered to R^2 *.737481*. However, the relative model-fit is higher, with a better score on the AIC, SC, and log-likelihood. The heteroskedasticity is again questionable. The KB-test and BP-test namely have a discrepancy in the significance of the heteroskedasticity. The Multicollinearity is slightly lowered by leaving out the *BINEIST* and *PTNEIST*. The SA-diagnostics still indicate strong positive autocorrelation, but highly insignificant (*.44121*), even more then in the OLS with all PUGI-variables (*.40979*).

In conclusion, from the neighbourhood PUGI, only the blue infrastructure has a significant relation with mental health problems, where in general more blue-infrastructure tends to lower the prevalence of mental health problems. After exclusion of the highly insignificant *PTNEIST* and *BINEIST*, the absolute model-fit was lower, but the relative fit was higher compared to the model that included all 4 PUGI variables. The SA-statistics did not show significant auto-correlation in the model-residuals.

5.4.8 Buffered Models: Mental Health Problems

Next up is modelling the PUGI on a *neighbourhood +400m* buffer in relation to mental health problems using an OLS-model *(appendix 18).* The OLS-model calculates the correlation between the *neighbourhood+400m* meter PUGI (*PT400ST; BI400ST; BL400ST; GR400ST*), the covariables (*FSI; LINC%*) and mental health problems (*MMHP%*). This *neighbourhood+400m* OLS-model has a slightly lower adjusted-R^2 (*.733680*) compared to the neighbourhood-model (*.735102*). The relative model-fit is also slightly worse when looking at the AIC, SC, and log-likelihood-tests. The PUGI-coefficients are still mostly insignificant, but in this model buffered green-infrastructure (*GR400ST*) is significant instead of non-buffered blue infrastructure (*BLNEIST*) in the neighbourhood model. Heteroskedasticity indicators are giving questionable results (non-significant KB-test in combination with a significant BP-test); however, the visualisation of the residuals this time *(figure 5-12)* suggests that heteroskedasticity is more likely than within the non-buffered OLS-model (*figure 5-11*). There is clear dispersion among the residuals when the % of people with mental health problems increases. SA is still insignificant (*global Moran's I*) and also the other SA-indicators are insignificant. Concluding on these diagnostics the *neighbourhood+400m* OLS-model suggests a positive

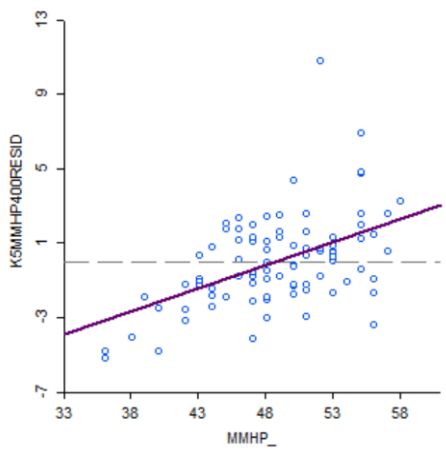


Figure 5-12: Heteroskedasticity in the residuals of the 400-meter buffer OLS

significant relation between green infrastructure (*GR400ST*) and the prevalence of mental health problems, meaning more PUGI within 400 meter of a neighbourhood increases mental health problems in Amsterdam neighbourhoods. This is a counter-intuitive outcome that is most likely caused by strong explanatory power from the covariables that over-rules the *GR400ST* coefficient.

The *neighbourhood+800m* buffer OLS-model (A*ppendix 19*) has a lower absolute model fit than the neighbourhood OLS-model, as well as the *neighbourhood+400* meter OLS-model with a adjusted R^2 of *.720377*. The relative model-fit is also lower than both other aggregation levels with a higher SC and AIC and a lower log-likelihood. All 4 PUGI-variables are insignificant; meaning the PUGI within 800-meter has no expected significant relation to mental health prevalence in neighbourhoods (high chance the null-hypotheses; there is no correlation, is true). Multicollinearity is still low but knowing the lack of significant explanatory power in the independent variables this model is not satisfactory and the relation using the PUGI on a *neighborhood+800m* level does not seem to be a great predictor of mental health.

Concludingly a OLS-model is within all 3 aggregation areas the best model-choice based upon the OLS SA-statistics. The trend in all three models is the importance of the two covariables (*FSI; LINC%*). The high significance (low-probability) of these covariables is seen on all 3 three aggregation levels and they thus have significant impact on the expected mental health problems prevalence. The SA-coefficients give no reason to believe a spatial lag-model or spatial error-model would give a better result as these are not significant. The high *adjusted* R^2 the OLS-models achieve (neighbourhood *.735102;* neighbourhood+400m *.733680;* neighbourhood+800m *.720377)* validate that the OLS-models perform satisfactorily.

5.4.9 OLS-Model: Loneliness (K5)

Loneliness among Amsterdam-residents is problem that affects around 50% of the population, and therefore it embodies a serious health-problem that can catalyse other mental and physical health problems such as depression, unhealthy food-patterns (obesity) and reduced social activities and safety-net. Modelling the relation between the *GRNEIST*, *BLNEIST*, *PTNEIST*, *BINEIST*, *FSI*, *LINC%* and loneliness (*LO%*) therefore can help to discover if the PUGI together with the covariables have a significant correlation with the prevalence of loneliness. The OLS-model *(Appendix 20)* is using K5 spatial weight to inspect the SA-indicators. The neighbourhood OLS-model indicators are shown in *table 5-10 (Next Page)*

Table 5-11: Model-diagnostics, OLS test with K5 spatial weights (loneliness %) *(Appendix 20)*

Test	*Value*	*Probability (ρ)*
Moran's I	4.9456	.00000
Lagrange Multiplier lag (Robust)	10.4020 (.2979)	.00126 (.58520)
Lagrange Multiplier error (Robust)	17.1316 (7.0275)	.00003 (.00803)

Test	*Value*	*Probability (ρ)*
Adjusted R^2	.634450	*N.A.*
F-statistic	27.0341	1.02138e-017
Log likelihood	-229.407	*N.A.*
Akaike info criterion	472.814	*N.A.*
Schwarz criterion	490.39	*N.A.*
Multicollinearity condition number	8.589815	*N.A.*
Breusch-Pagan	22.9174	.00082
Koenker-Bassett	7.5504	.27293

Variable	*Coefficient (β)*	*Std. Error (σ)*	*T-statistic*	*Probability (ρ)*
CONSTANT	39.9443	1.34864	29.6181	.00000
BLNEIST	.0392382	.408185	.0961285	.92366
PTNEIST	-.594676	.399324	-1.4892	.14018
BINEIST	-.158586	.434645	-.364864	.71613
GRNEIST	-.128438	.371156	-.34605	.73017
LINC%	.882829	.0775878	11.3784	.00000
FSI	-3.60421	.913822	-3.9441	.00017

The diagnostics show a satisfactory, but relatively low, absolute model fit (*R^2 .634450*) compared to the previous OLS-models on the other dependent health-variables (*BMI30; MMHP%*). Leading to the first conclusion that the independent variables have a stronger absolute model fit (adjusted R-squared with the neighbourhood OLS-model) when modelling mental health problems (*.735102*) and obesity (*.663955*) compared to loneliness (*.634450*). The relative model fit is additionally lower than the previous models, the AIC and SC-tests are higher (*472.814; 490.39*) and the log-likelihood is lower (*-229.407*). The constant-coefficient is highly significant (*.00000*). The variable-coefficients imply that a higher FSI in general is negatively significantly corelated with loneliness, higher densities thus relate to lower loneliness prevalence. That seems logical as higher densities often mean apartment-housing with many nearby amenities that catalyse social-life and a higher population density. Low income people (*LINC%*) has the reverse effect, and it shows that lower SES-people tend to be significantly positively correlated to more loneliness prevalence. The PUGI-variables are all insignificant in relation to loneliness (*PTNEIST .14018; BINEIST .71613; GRNEIST .73017; BLNEIST .92366.*)

The multicollinearity is well below the threshold (*<30*) with *8.58981,* thus overlapping correlation between variables is at acceptable levels. The heteroskedasticity indicators (BP-test and KB-test) have different outcomes.

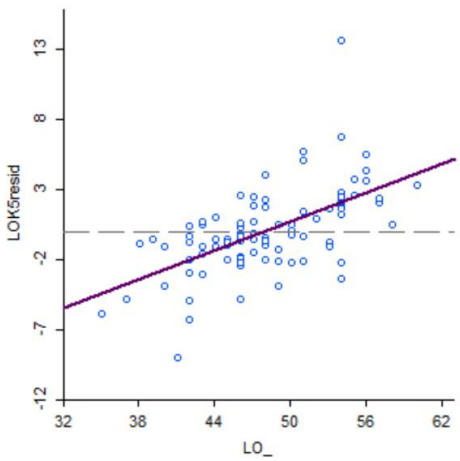


Figure 5-13: potential heteroskedasticity in the OLS-residuals (loneliness)

Whereas the BP-test implies significant heteroskedasticity, the KB-test does not. In *figure 5-13 (previous page)* the (lagged) residuals show that heteroskedasticity is not explicitly present as the residuals vary similarly among all loneliness prevalence levels.

The SA-statistics show highly significant positive SA within the residuals (*global Moran's I .00000*). The (Robust) LaGrange multiplier lag and error diagnostics imply a spatial-error model is the optimal model as both the normal and the robust LaGrange multiplier error-diagnostics are significant, whereas this is not the case with the Lagrange multiplier lag-diagnostics. In conclusion, a spatial-error model will be made in addition to the OLS-model as this may improve the model-diagnostics and fit.

5.4.10 Spatial Error-Model: Loneliness (Q1)

The spatial-error model (*appendix 21*) makes use of Q1 weights as they require symmetric weights just like the spatial-lag model. The spatial-error model outcomes greatly improve the absolute model fit (R^2 *.724282*). Up from *.634450* with the OLS-model. The relative model-fit is also improved, with both the AIC and SC tests being lower and the log-likelihood being higher. These first model-fit diagnostics thus seem to confirm the spatial-error model outperforming the OLS-model. To reduce the residuals and SA-related errors the spatial-error model introduces a new variable: *LAMBDA (Λ)*. The LAMBDA-coefficient represents the spatially correlated errors and is highly significant (*.00001*). The covariables *FSI* and *LINC%* are still highly significant (*.01470; .00000*) and additionally the constant-coefficient is highly significant (*.00000*). The PUGI-variables became more insignificant in the error model, thus the relation between PUGI and loneliness stays disputable.

Heteroskedasticity is this time significant with the BP-test being highly significant (*.00002*). The error-model residuals-scatterplot in *figure 5-14* is similar to the OLS-model residual scatterplot in *figure 5-13 (previous page).* Overall, the same level of heteroskedasticity thus is seen in the spatial-error model. In contrast to the OLS-model diagnostics the spatial-error model shows the *likelihood ratio test* is still significant, meaning there is still spatial dependence/autocorrelation in the residuals.

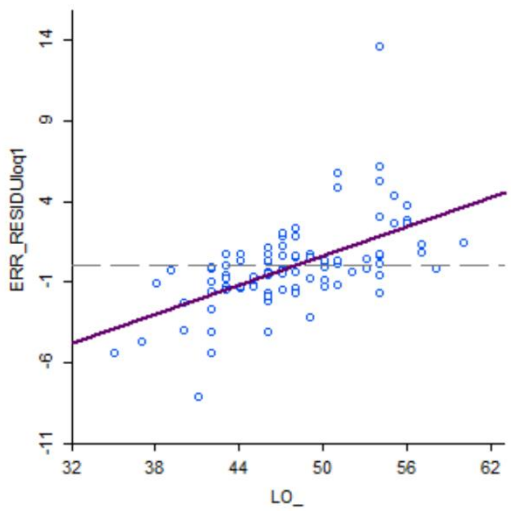


Figure 5-14: residual scatterplot of loneliness OLS-model

Concludingly the spatial-error model has a better absolute as well as relative model-fit compared to the OLS-model on a neighbourhood aggregation level. The error-model shows the PUGI-variables are not significantly correlated to the loneliness prevalence. Furthermore, it shows the model-residuals showcase heteroskedasticity, meaning the errors are not spread out evenly over the different levels of loneliness-percentages.

5.4.11 Buffered Models: loneliness

The neighbourhood aggregation-models show that the PUGI does not have a statistically significant (*probability* <.05) relation with loneliness, the *neighbourhood 400m* buffer and *neighbourhood 800m* buffer are now modelled to see if the PUGI within these proximities is significantly correlating to the loneliness-prevalence.

The OLS-model with the *neighbourhood+400m* PUGI-variables (*appendix 22*) shows a lower absolute model-fit compared to the neighbourhood model (R^2 .632000). All 3 of the relative model-fit indicators worsened with the addition of the 400-meter buffer. The coefficients their probabilities are slightly better on average but again none of the PUGI-variables are significant ($p < .05$), whereas the FSI (.0006) and LINC% (.00000) correlate highly significant with the loneliness-prevalence. There is less multicollinearity (8.106766) and again disputable heteroskedasticity, with the BP-test being significant (.00160) and the KB-test being insignificant (.30493). The SA-diagnostics suggest there is positive SA and that an spatial-error model is supposedly a better fit with both the normal and robust LaGrange multiplier error diagnostics being significant (.00050; .00000).

Indeed, the *neighborhood+400m* buffer spatial-error model (*appendix 23*) has a higher absolute model fit (R^2 .729417), compared to .632000 in the OLS-model. The relative model-fit diagnostics have worsened, meaning the OLS-model has a better balance between simplicity and explanatory power (R^2). The significance of the 4 PUGI-coefficients is still above the significance-threshold of .05. The LAMBDA-coefficient (spatial-error coefficient) is highly significant (.00001) and both the *CONSTANT*, *FSI* and *LINC%* coefficients are also highly significant (.00000; .01154; .00000).

The *neighbourhood+800m* buffer OLS-model (*appendix 24*) showcases similar results in the diagnostics as the n*eigbourhood+400m* buffer OLS-model. The absolute model fit is similar (R^2 .633471), and the relative model-fit indicators have similar values compared to the neighbourhood and neihgbourhood+400m buffer OLS-models. The PUGI-coefficients are not significant, only the covariables (*FSI* .00002; *LINC%* .00000) and the *CONSTANT* (.00000) are. Multicollinearity is lower, this is a trend observed in alle larger-aggregation level models in this analysis. The more PUGI-buffer is included the less they correlate with each other, the covariables and dependent variables.

The BP-test is highly significant (.00307), while the KB-test is insignificant (.035294). Heteroskedasticity indicators thus give no clear proof of this phenomena, and the scatterplots presented previously show why; the residuals are dispersed, but not per se more in a certain direction on the x-axis. The *neighbourhood+800m* buffer has significant positive SA, again the LaGrange multiplier error-indicators are most significant, preferring an error-model over a lag-model.

The spatial-error model of the *neighbourhood+800m* buffer PUGI in relation to loneliness prevalence (*Appendix 25)* has a higher absolute model fit (R^2 .732203). The relative model-fit indicators in addition have also improved, with both the AIC and SC tests as well as the log-likelihood indicators being improved

compared to the OLS-model. The coefficients of the variables are more significant, with the public transport variables (*PT800ST*) being significant this time (*.03609*), while the other PUGI-variables stay insignificant in relation to loneliness. The *CONSTANT*, *FSI* and *LINC%* coefficients are highly significant, and in addition the LAMBDA-coefficient is highly significant (*.00000*). The likelihood ratio test is highly significant (*.00003*); meaning the SA is still present, even with the introduction of the LAMBDA-coefficient in the spatial-error model.

Concludingly the spatial-error model improved the model-fit compared to the OLS-model; the significance of the *PT800ST* is striking, as this variable has a strong coefficient in this spatial-error model (*1.23619*) that is also significant (*.03609*). This would suggest that having more public transport-stops within 800 meter from your neighbourhood is positively correlated with more loneliness, a counter-instinctive finding. Therefore, this is most likely a coincidence (even though the significant probability says otherwise).

5.4.12 Multiple-Regression Summary

The 3 health variables required different model-types to ensure best-fit. These model-types and the hereof coming model-fit indicators are compared and explained in this chapter. Within the obesity models (*BMI30*) the normal and robust Lagrange-multiplier (lag) indicators were systematically significant, indicating that the SA in the residuals required a lag-coefficient for better fit. The spatial-lag models indeed outperformed the OLS-models with every aggregation, with an increase in the absolute and relative model-fit indicators and a highly significant lag-coefficient (*W_BMI30*). With the Mental health problem-models (*MMHP%*) the SA-statistics on the other hand were insignificant. Suggesting that the OLS-model is the best option. Indeed, the model-indicators are already showing acceptable model-fit with the OLS. The loneliness-models (*LO%*) had significant Lagrange-multiplier error-indicators (normal as well as robust); and thus, a spatial-error model was used to analyse the correlations this time. The *LAMBDA*-coefficient (error-coefficient) indeed increased the absolute and relative model fit.

Both 3 independent health-variable-models had highly significant *CONSTANT* (OLS, error, lag), *LAMBDA* (error) and *W* (lag) coefficients (p *.00000*) and in generally highly significant *FSI* and *LINC%*-coefficients ($p < .05$). The PUGI (independent) variables, green, blue, cycle and public transport-infrastructure were in general (highly) insignificant. There were a 2 exceptions; the BMI30-model without neighbourhood-buffer indicated a significant correlation with public transport infrastructure (*PTNEIST*); in the OLS-model as well as the spatial-lag model. Additionally, the mental health problem-model (*MMHP%*) using a *neighbourhood+400m* buffer had significant correlation with green-infrastructure (*GR400ST*). Because of the scarcity in which these independent variables were significant ($p <.05$), the overall conclusion about the multiple-regression is that these variable have little to no impact on the dependent health-variables, on all aggregation-levels. The different models and variables that are used and their model-indicators are compared in *table 5-12 (next page).* The **bold** R-squared-values show the model with the highest fit within a specific health-variable.

Table 5-12: Comparison of Models

Dep. Var.	*Significant. Indep. Variables (p <.05)*	*Buffer*	*Model (App.)*	*(Adj.) R-Squared*
BMI30	*CONSTANT; FSI; LINC; PTNEIST*	*+0*	*OLS (1.0)*	*.663955*
	CONSTANT; FSI; LINC; PTNEIST	*+0*	*OLS (1.1)*	*.667163*
	W_BMI; CONSTANT; FSI; LINC; PTNEIST	*+0*	*LAG (1.2)*	***.789065***
	CONSTANT; FSI; LINC	*+400*	*OLS (2.0)*	*.626090*
	W_BMI; CONSTANT; FSI; LINC	*+400*	*LAG (2.1)*	*.762863*
	CONSTANT; FSI; LINC	*+800*	*OLS (3.0)*	*.628083*
	W_BMI; CONSTANT; FSI; LINC	*+800*	*LAG (3.1)*	*.763369*
MMHP%	*CONSTANT; FSI; LINC*	*+0*	*OLS (4.0)*	*.735102*
	CONSTANT; FSI; LINC	*+0*	*OLS (4.1)*	***.737481***
	CONSTANT; FSI; LINC; GR400ST	*+400*	*OLS (5.0)*	*.733680*
	CONSTANT; FSI; LINC	*+800*	*OLS (6.0)*	*.720377*
LO%	*CONSTANT; FSI; LINC*	*+0*	*OLS (7.0)*	*.634450*
	LAMBDA; CONSTANT; FSI; LINC	*+0*	*ERROR (7.1)*	*.724282*
	CONSTANT; FSI; LINC	*+400*	*OLS (8.0)*	*.632000*
	LAMBDA; CONSTANT; FSI; LINC	*+400*	*ERROR (8.1)*	*.729417*
	CONSTANT; FSI: LINC	*+800*	*OLS (9.0)*	*.633471*
	LAMBDA; CONSTANT; FSI; LINC	*+800*	*ERROR (9.1)*	***.732203***

5.5 Research Question 4

5.5.1 Introduction

The planned 2050 PUGI in Amsterdam is based upon a study called *Groenvisie 2050.* This *green-vision 2050* was commissioned by the Amsterdam city-council. This long-term policy document emphasizes the importance of PUGI for Amsterdam citizens in many social and health related domains and tries to implement these domains in actual policy-implementations. Emphasis is put on the value of nature as a place for relaxing, sporting, and playing for all ages and social-groups and climate-change resilience. Laurens Ivens, the current municipal green-councillor mentions the following in the preface of the document: "*Green infrastructure is important for a pleasant and healthy living environment and is needed to protect the city against drought, heat and extreme rainfall. Greenery is of vital importance to Amsterdam. In a growing city where there is an increasing demand for housing and the public space is our backyard; attention needed for the construction, preservation and management of greenery, so that everyone can continue to enjoy it*". (Wijten, 2020, p.5). See *appendix 8* for the original Dutch quote. This quote is supported by a infographic from the *Groenvisie 2050* (*figure 5-4*).

Figure 5-4: The 4 focal points which the 2050-PUGi aims to support and atrengthen

5.5.2 2050 Policy Implementations

The 2050 PUGI-vision is exclusively looking at the municipality of Amsterdam, the same scope as this study, and therefore the multiple regression and theoretical-findings can be used to support or criticise the 2050-vision. The multiple-regression outcomes are reflected to the 2050-policy to make a substantiated argumentation regarding the expected effectiveness of the 2050-PUGI. The 2050 green-vision proposes some idealistic as well as realistic and practical spatial policy implementations that help improving the 4 themes from *figure 5-4 (previous page).* These vary from non-spatial implementations (a concierge for every park) to changing governance-structures such as increased municipal-citizen partnerships for PUGI-maintenance and social-inclusion. But the 2050 PUGI-vision also prominently includes *physical* PUGI investments in additional parks/forests, green-connections/corridors, and *grey-area* transformations. A resume of the vision its policy-implementations is seen in *table 5-13.*

Table 5-13: the policy-pillars of the 2050 green vision and their elaboration

2050 target	Implementation and elaboration
Within 10-minutes (walking) a park for every citizen	New parks will cover op the black-spots that currently lack a park within 10-minutes walking distance.
Construction and maintenance of PUGI becomes a partnership between the municipality and citizens	Cooperation of the PUGI-maintenance and planning ensures the PUGI is integrated in the social-structure of the neighbourhood. E.g., people maintain small gardens in the PUGI that they harvest for food or flowers. The to-be park-concierges will help maintain this relation.
Within 15-minutes (cycling) rural-green can be accessed (e.g., Twiske, Amsterdamse Bos)	The addition of a new urban forest within the municipal borders (similar to the Amsterdamse Bos). Introduction of green corridors, quick, save and lush walking and cycling routes that connect PUGI-hotspots for quicker and more comfortable access.
Green, unless… policy	All Gray-infrastructure (e.g., sidewalks, parking spaces, roundabouts) is turned into PUGI, unless the presence of the asphalt/pavement is crucial for the usability of the areas.
Non-public urban green infrastructure will become public	Public allotment gardens, sport-facilities (e.g., running tracks, football pitches, school gardens) are transformed from private-infrastructure into PUGI for increased coverage and diversity in the PUGI-offering.
Every park gets a concierge	Parks get a concierge that maintains extra facilities like toilets, helps in planning sport and social-events, improves the feeling of safety and functions as point of contact in miscellaneous situations.
Users determine park-amenities	Less monotone parks, every park gets amenities and facilities that fit the neighbourhood/surrounding area and the users of the park. Family - neighbourhoods get different park-layouts and amenities then student or senior neighbourhoods for example.
New parks get fitted in the existing urban fabric	New parks will be included in the existing urban fabric by combining existing smaller patches of PUGI, turning wastelands in PUGI and instant future implementation in large-scale urban (housing) developments.
Parks get more focus on physical activity	More nearby PUGI means companies and educations-institutions can motivate their employees and student to get involved in physical activity in lunch-breaks for example. Sport-facilities get include in the PUGI; making more people be able to use more professional sporting facilities. IN addition, parks get dedicated sport-facilities in areas where there is need for this type of amenity.

5.5.3 2050 PUGI Strongpoints

Make private-UGI more accessible. This focal-point is substantiated by the multiple regression-analysis that repeatedly showed the negative (and highly significant) impact a lower-SES has on the health variables (on a neighbourhood level). Part of the focus of the 2050-PUGI lies on the importance of *social wellbeing catalysator* as an intrinsic characteristic of PUGI. Actual 2050-policy that supports this ambition is also in place. For example, *private*-UGI is to be transformed into *public*-UGI, increasing (financial) accessibility for people that are currently not able to use this high-quality and amenity-rich facilities. Non-planned usage becomes more accessible when no fee, memberships or reservation is required. For example, municipal football-pitches, tennis-courts, and running-tracks become free-to-use. Opening these currently private facilities for free-usage will increase the overall amount of PUGI that can be used, but more important the facilities allow and tempt users to initiate physical activity more than existing (busy) urban parks. These, often large facilities, are located mainly in the outskirts of the city and therefore facilitate mainly the people that currently have the highest obesity, mental health problem and loneliness prevalence. Additionally, private green from allotment-gardens (Dutch: Volkstuinen), are to be partly transformed from private green infrastructure into PUGI, further increasing the public at the expense of the private infrastructure.

Increase social-inclusion and cohesion in PUGI. Neighbourhood-initiatives that take place within the 2050 PUGI will be better supported to turn PUGI into clusters/hot-spots of community-activities. This can strengthen the community-feeling in Amsterdam. In addition, low-quality green, that is mostly present in low-income areas, will get upgraded (in consultation with the users and neighbours) to add value to the neighbourhood and be more inviting to use. The concierges that all parks will supposedly get in 2050 additional help improve safety and monitor the quality and usage of the PUGI. The community-role PUGI currently already has (e.g., birthdays/barbecues, team sporting, theatre/music activities), will be increased by including citizens in the proximity of the PUGI to help in the decision making process, adding new amenities and activities that fit the socio-cultural visitor-profile. These amenities can be of a wide variety: playgrounds, urban-farming initiatives, and cultural facilities like a public-podium.

Increase PUGI-accessibility and proximity. The 2050-vision aims to make PUGI more accessible, opening up the parks, waters and urban-forests to neighbourhoods using so called *green-corridors,* concatenation of small PUGI-patches into a larger PUGI-cluster, improving intra-accessibility between PUGI. Additionally, the 2050 PUGI-vision aims to incorporate bike, walking and public transport infrastructure in the PUGI, to help increase accessibility to the (potential) users. This may improve the accessibility to qualitative PUGI for residents that currently lack access to PUGI-facilities due to reduced mobility. This in return may help reduce the current discrepancy between high and-low income people regarding the %-usage of PUGI. Making PUGI inclusive social-hotspots and simultaneously improve physical park-accessibility can additionally improve the loneliness prevalence. Additional incremental and local PUGI additions, due to *grey-area* transformation (such as green facades, roofs, roundabouts, and parking-spaces), help increase PUGI-accessibility by bringing

the PUGI in the direct proximity of people, reducing the travel-time to PUGI even more. The proposed PUGI-transformation of local *grey-areas* increases climate resilience while simultaneously having a positive effect on miscellaneous mental health problems, as seen in the literature.

Increase the space for physical activity at the expense of *grey-infrastructure*. The municipality found that usage of PUGI for physical activity decreased slightly between 2013 and 2018 (OIS, 2018) shows that. A trend that was caused by the over-crowding of some parks. By creating (new) parks, connect small PUGI-patches into coherent green and water-corridors/clusters and transforming sport facilities and allotments gardens into public terrain the absolute availability of PUGI is increased. Ideally reducing the usage-intensity where needed. Example of current developments that go hand-in-hand with the 2050 PUGI-policy on the short-term is the tunnelling of the A9-highway that runs through the Bijlmer-neighbourhood (*figure 5-6)*. The tunnelling-phase was completed recently (2020) and will in the near-future (2021-2022) be covered by a modern park (*figure 5-5)*. The park-infrastructure includes extensive blue-infrastructure and high-quality bike-infrastructure. For example 4 large cycle-bridges will connect the new PUGI with the existing (PUGI) infrastructure (Rijkswaterstaat, 2020). This park however is located outside of the central city; therefore, additional PUGI additions within more central areas are required to relieve the usage-intensity in these areas, there is however little space there for large-scale additions such as this project. The transformation of *grey-areas*, upgrading of blue-infrastructure and merging of small existing PUGI-patches is a more viable option in these locations.

Figure 5-5: Current tunnelling construction-site of the A9-highway that will get PUGI on top after completion (Rijkswaterstaat, 2020)

Increase the particulate matter retention. By means of an new (large) urban-forest and the strategic positioning of trees along busy roads particulate matter can be filtered and/or averted from the residents that live in polluted areas or on polluted arterials. Car-arteries like the *Prins Hendrikkade*, *Overtoom* and *Stadhouderskade* namely still exceed several

Figure 5-6: Artist impression of the new park on top of the tunnel (Rijkswaterstaat, 2020)

air-pollution thresholds (Andreas Mack & Duyzer, 2019). The tunnelling of the A9 *(figure 5-5; figure 5-6)*, therefore is also a great example in diverting the direct particulate matter emissions in the Bijlmer-area. But also the proposed speed-limit downgrade from the A10-west from 80 km/u to 50 km/u will lead to significant local and regional particulate matter reductions (Koops, 2021). The actual filtering of particulate matter from the air by means of small parks and road-side PUGI is minimal, studies have shown. However, significant particulate matter filtering does take place in larger metropolitan-scale PUGI-areas at the edge of the city. Therefore, the addition of a new urban forest can positively impact the overall particulate matter retention in the whole Amsterdam-region significantly. Particulate matter on some locations in Amsterdam reduces life-expectancy with 13 months, and on average accounts for 75% of the external factors that cause work-related sick-leave (Rijksinstituut voor Volksgezondheid en Milieu, 2018). Thus, the addition of a large new PUGI-area (in the form of a forest) that has sufficient air filtering-ability will increase air-quality while decreasing thereof coming mortality and sick-leave.

Create mentally healthy neighbourhoods. The 2050 PUGI-vision emphasizes on incorporating PUGI in as much *grey-areas* as possible. The incorporation of PUGI in all scale-levels, from green facades to a new urban forest, helps in creating stress free environments. Local small-scale green like greening of the pavement and walls in theory reduces stress by being visually pleasing then grey-infrastructure. Studies have shown that the visibility of green reduces many stress-hormones and increases concentration and work-productivity significantly. The local as well as regional PUGI-developments that are pursued in the 2050-policy thus in theory improve the mental health domain in neighbourhoods. In addition, local PUGI decreases the urban-heat island effect. With that, reducing productivity-problems and miscellaneous mental (stress and sleep problems) as well as physical (cardiovascular diseases) health-problems can be reduced (Dinther & Weijers, 2016; Milgen, 2016). A near-future example of hyper-local urban-PUGI integration is seen in *figure 5-7* (Muncipality of Amsterdam, 2021), the *Valley* apartment-complex facilitate a stress-reducing, lush and cool living environment. Expected to be completed in summer 2021 (Muncipality of Amsterdam, 2021).

Figure 5-7: PUGI-integrated with apartments in the Zuidas

Incorporate green more significantly in newly developing urban areas. Strong integration of PUGI in the urban fabric (existing and new) ensures an increase in PUGI does not reduce the density of the urban fabric (FSI) and makes the urban PUGI more accessible and the surrounding neighbourhoods more dense, walkable/cyclable, and climate-resilient. The incorporation of new PUGI within the existing urban fabric reduces the need for far-away PUGI-areas at the city-edges. Inner-city parks, with the inclusion of a concierge

and well-used neighbourhood-amenities will have improved social-control and lower travel-distances for users, reducing the visiting-threshold. In addition, by making the PUGI an integral and accessible part of the neighbourhood, the municipal-citizen collaboration regarding the maintenance will be more accepted and efficient. Example of strong PUGI-integration in a large residential development is the *Sluisbuurt* in the Zeeburg. The Sluisbuurt will be finished incrementally within the next 10 to 15 years according to the current pace of development. The area includes 5.500 climate-neutral housing-units, a large waterfront-park with a public-pool and car-free boulevards/ squares with canals and green-corridors. These design-features are aligning with the 2050 PUGI-vision. In *figure 5-8* the strong integration of PUGI, infrastructure and other amenities is visualized (BOOM landscaping, 2021). The Sluisbuurt concludingly is a great example neighbourhood-design that follows the 2050 PUGI-vision.

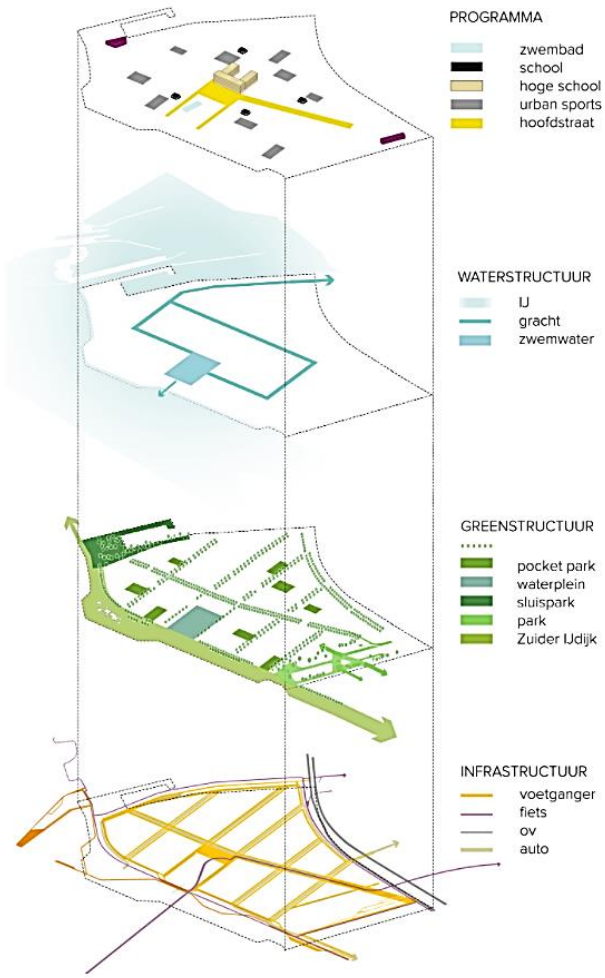


Figure 5-8: The PUGI, (public) amenities and infrastructure-design of the Sluisbuurt

5.5.4 2050 PUGI Weak points

The multiple-regression analysis brought to surface that the absolute amount of PUGI, in general, is not a significant ($p < .05$) contributor to obesity, mental health problems and loneliness. The ambitions of the municipality to make PUGI a catalysator for health improvements therefore is not viable without reducing broader socio-economic problems and reduce the SES-gap between neighbourhoods. The 2050-policy aims to incorporate the ideas and efforts of citizens that life in the proximity of new PUGI for the PUGI development. However, people that already possess a strong social-network and social-capital are more likely to join the participation-process and thus volunteer in PUGI planning, maintenance and activities (Kleemans, 2014). Therefore, low-SES people, that currently make the least use of PUGI and have most health-problems, are the least likely to be represented in the planning-process. A strong social agenda and inclusion-efforts should accompany the PUGI-vision to ensure the health-improvements are fairly distributed over all SES's. In addition, the 2050 PUGI-policy must be aware it uses a strong pragmatic approach in achieving its social, health and climate-related goals. Overly expensive and prestige-like PUGI-interventions

must be avoided to ensure the PUGI-funds have a strong return of investment in these domains. Example of a recent wasteful PUGI-project was the placement of 8 so-called *City-Trees (figure 5-9)* alongside several car-arteries in Amsterdam (Green City Solutions, 2021). The development of these trees was partly funded by the European Union back in 2014. The City-Trees, according to their inventor, would reduce local PM_{10}-values with 19% and have the same 'filtering' capacity of 275 actual trees. The substantiation of these claims was highly questionable from the start of the project. The City-Trees indeed, did not lead to a significant reduction of PM_{10}, and even slightly increased the local NO_2-values and therefore were removed after just 8 months. Therefore, the high €200.000 initial investment led to public critique in the direction of the municipality. To prevent future questionable investments, a strong pragmatic control-system must accompany the 2050 PUGI-policy. The ambitious and idealistic ideas that are in the 2050 PUGI-vision must be tested robustly beforehand on effectiveness regarding (among other things) public health improvements.

Figure 5-9: City Tree along a busy Berlin-road

Finally, establishing such a long-term PUGI-vision comes with inherent continuity-risks due to an ever-changing political landscape in Amsterdam. Securing the required political-will and funding for the next 29 years is therefore not guaranteed. But even with a predominantly *green* city council there is ongoing critique on the municipalities current PUGI-policy, that shows some discrepancies with pursuing their own 2050 PUGI-vision. For example, large public trees are removed at busy several streets to accommodate remodelling and maintenance. This is seen in several inner-city car-arteries and squares. Examples are the *Rozengracht* and *Frederik Hendrikstraat*, both located in the PUGI-scarce late 19th-century western part of the city. The removal of large trees increases the urban-heat island effect that in return increases physical health-problems during heat-stress moments. Small trees will be planted after completion of street-design changes, but it will take many decades before they have the same health-benefits as the removed trees. Therefore, sustaining credibility in pursuing ambitions PUGI-policy is much needed.

5.5.5 Conclusion

The ambitious 2050 PUGI-vision is a solid and well substantiated starting-point for improving the accessibility, usability, and health-impact of the current PUGI. Many of the policies are expected to improve overall health and increase physical activity and social life that takes place in the PUGI, however mostly based upon theoretical evidence. The PUGI-vision focusses a lot on citizen participation, which is a practice not all people take part in equally, especially the people that are currently in bad physical state, have mental problems or are stressed and socially excluded (lonely) may be overlooked. Therefore, the participation as a result may only increase amenities and activities that cover the needs of the wealthy, social and fit people. Actively

including people from all socio-economic classes is a municipal task they should take seriously regarding the PUGI-developments. In addition, the 2050 PUGI-vision should stay pragmatic in its execution to maintain broad political and societal support in its funding and execution. Finally, the current Green Left led municipal council must aim to guarantee the continuation (of funding and political support) of the ambitious 2050 PUGI-policy. Efforts should be made to establish a broad and sustainable political base regarding PUGI-investments to make it less prone to changes in the political discourse.

6. Conclusion

The importance of PUGI in relation to health is studied in many geographical areas worldwide, the results from these studies varied due to locational characteristics, cultural differences, and historical developments. This case-study of Amsterdam is no different in this aspect. The main research question, '*The case of Amsterdam neighbourhoods: what public urban green infrastructure variables impact health variables, and how can future green infrastructure interventions change this relationship?*', was dissected in 4 research question that together made up a chronological pathway for the answering of it. The hypothesis was that within a multivariate spatial regression-model, more PUGI does only slightly improve health variables in Amsterdam. It is expected that important covariables, like the socio-economic status and urban density, have a larger and more significant impact on health variables (compared to PUGI). This hypothesis was based upon the knowledge that the PUGI in Amsterdam is concentrated in the outskirts where the density and income (2 theoretically important covariables) are generally low compared to inner-city parts.

The vast theoretical framework that is already in place on this topic gave way for selection of variables that are deemed to be of interest. This selection was well-substantiated using a combined theoretical body of case-studies as well as global (WHO) reports. The variables that were of interest mostly are available at one of the open geo-portals in the Netherlands. These portals provide access to up-to-date official governmental datasets on PUGI and health related topics. Additionally, some important covariables are available on these portals. However, some variables, dependent as well as independent, were not available on the required neighbourhood aggregation level. Example of this are the fast-food infrastructure (restricted data-set from the Chamber of Commerce) and the prevalence of cardiovascular diseases (not available on a neighbourhood aggregation due to medical-sensitiveness of data).

The multiple regression showed that the relation between PUGI and the health variables (obesity, mental health problems and loneliness) was generally not significant. Also, when aggregated to larger areas using buffers, the relation between PUGI and the 3 health variables was generally not significant. The expectation was that the buffers would reduce the MAUP, and in general this improves model performance; however, this was not the case. The covariables, FSI and percentage low-income people (as a measure for SES), were systematically significant variables in the models. The FSI and SES-covariables were correlating strongly significant with all of the three health-variables, with both having strong-coefficients and a highly significant probability of (close to) $.00000$. These findings are only relevant on a neighbourhood aggregation level, the correlations and their significance say nothing about the whole city of Amsterdam, as that would be an ecological fallacy.

While bike-infrastructure is in none of the regression-models a significant explanatory variable in relation to the health-variables, the importance of this infrastructure is substantiated strongly in other studies. The implementation of bike-infrastructure in this model is hard, as the quantity of bike-infrastructure does not well reflect the usage and attractiveness of the infrastructure. This variable therefore is statistically

insignificant, but theoretically a significant contributor to the improvement of health variables. The same holds for the blue and green infrastructure. The correlation with health-variables on and neighbourhood level is not statistically proven, however the theoretical substantiation on the positive effects it has on overall citizen health make believe that the PUGI has urban-wide benefits, that are just not significant on the neighbourhood aggregation level. The neighbourhood aggregation was an experimental aggregation level based upon the assumption that it was the ideal unit-of-analysis for policy makers. Most studies that found significant corelation between PUGI and health used larger units of analysis or compared several socio-economic similar neighbourhoods with each other to intrinsically account for the covariables.

The 2050 PUGI-vision included ambitious changes to the public space, that in return potentially could impact the health of the Amsterdam citizens. The changes in the PUGI that are pursued in 2050 were analysed on the potential effectiveness, with the knowledge from the multiple-regression outcomes and existing theory on this subject. The overall outcome is that for a large part the 2050-vision reduces several health problems. The 2050 PUGI-vision does this by improving the way PUGI fits in the built-environment and make the functionality and amenities more conforming its local users. In addition, the 2050 PUGI-vision improves social-cohesion in the PUGI and surrounding neighbourhoods by creating community gardens, open existing sport facilities to the public (who are mainly located in the low-SES areas), introduce a point of contact (concierge), and include residents in the planning process.

Potential risk of the 2050 PUGI-vision may be that it relies upon citizen participation. While it can be expected that certain citizens will be more active in these newly added PUGI-functions. The people that currently lack ability to participate in social activities and sport activities may be left out of the potential health-benefits. Policy aimed in including these vulnerable groups must therefore ideally be integrated in the policy. The social gap among (potential) PUGI-users should be muffled; for example, by making use of local institutions in the planning process that have a strong foothold in the surrounding communities. Only then the ambitions that are envisioned in the 2050 PUGI-vision can grasp the most vulnerable people to help increase health-benefits among them.

The answer to the main research question '*The case of Amsterdam neighbourhoods: what public urban green infrastructure variables impact health variables, and how can future green infrastructure interventions change this relationship?*' is not straight-forwards due to the used PUGI-variables being insignificant in many of the models. There is no strong and significant relation between PUGI and the health variables on a neighbourhood-level in Amsterdam. However, the strong and statistically significant correlation between SES/FSI and the health variables implies that these factors are important additional policy-considerations. Regarding 2050, the inherent uncertainty of the eventual design of the 2050 PUGI makes the assessment of this policy prone to potential changes in this policy. However, based on the assumption the 2050 PUGI-policy will be fully carried out, significant health improvements are expected.

7. Recommendations and discussion

This study showed some model-limitations and statistical-problems that are preventable with the inclusion of different variables, use of smaller aggregation-levels and additional modelling techniques. The choices made in this study are based upon the available recent (2016-2021) open data that covers the case study-area with a neighbourhood-precision. For other areas and time-frames (or due to future data-availability improvements), more precise socio-economic and health data may be available to strengthen the statistical robustness of the models.

Additionally, the neighbourhood is not an ideal aggregation-unit due to irregular topology, size-differences and PUGI-concentrations that increase the MAUP-implications. A fine-grained regularly tessellated (grid/raster) surface such as the CBS-vierkanten are more structured in their topology while also increasing the sample size significantly. Together increasing the normality of the variables and the neighbourhood-matrix robustness and equality. However, most of the socio-economic and health data is currently unavailable on this aggregation-level, explaining the decision to use the neighbourhood aggregation in this study.

The dependent-variables that ideally would be included, such as cardiovascular-deceases and mortality (life expectancy) are only available on a district and municipality level, aggregating it to the (smaller) neighbourhood-level would dramatically reduce precision and correctness of the data. Some additional independent variables are currently available, but not as open data, such as the fast food-locations (Kamer van Koophandel register with pay-wall), an therefor excluded from the study.

Diversification of the PUGI-variables, where the quality and amenities of the PUGI are determinative for a certain PUGI quality-score/attractiveness-score would add additional quality to the model. This quality-score then weights into the impact certain PUGI has on the surroundings, where higher quality PUGI has a larger weighting then low-quality PUGI. The PUGI could in addition get a connectiveness-score that indicates how many people are within a certain distance or travel-time from the PUGI, giving central and well-connected PUGI additional model-impact.

Finally, the use of a geographically weighted regression-model (GWR) can be a solution to account for the significant SA-statistics and improve model-fit and reduce model-residuals and heteroskedasticity. The SA-statistics and heteroskedasticity-indicators in the variables are namely generally high and significant. The use of 1 regression coefficient (such as the OLS-model, lag-model, and error-model) has inherent limitations in its explanatory power for such diverse and/or clustered spatial data. GWR-modelling options were not available in the software that was used in this study (GeoDa) but could be implemented when using this data in more capable statistical-software.

8. Bibliography

Andreas Mack, & Duyzer, J. (2019). The Impact of City Trees on air quality in the Valkenburgerstraat (Amsterdam). In *TNO*.

BOOM landscaping. (2021). Sluisbuurt Amsterdam. Retrieved February 26, 2021, from https://boomlandscape.nl/work/sluisbuurt/

Boon, J., Westerink, P., Noten, I., van der Schee, G., & Derksen, E. S. (2020). *Gezonde Stad Index 2020*.

Bos, L. ., Horjus, B., Visser, S. L. ., & de Wilde de Ligny, S. C. . (2020). *Stedenbouwkundige ingrepen en de relatie met gezondheid en welzijn*. Deventer.

CBS. (2020a). Kencijfers wijken en buurten 2020. Retrieved February 28, 2021, from https://opendata.cbs.nl/statline/#/CBS/nl/

CBS. (2020b). Stedelijkheid (van een gebied). Retrieved February 28, 2021, from https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/stedelijkheid--van-een-gebied--

Chen, B., & Kan, H. (2008). Air pollution and population health: A global challenge. *Environmental Health and Preventive Medicine*, *13*(2), 94–101. https://doi.org/10.1007/s12199-007-0018-5

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., … Harris, P. (2020). The gwr route map: A guide to the informed application of geographically weighted regression. *ArXiv*, (April).

Congdon, P. (2019). Obesity and urban environments. *International Journal of Environmental Research and Public Health*, *16*(3). https://doi.org/10.3390/ijerph16030464

David O'Sullivan; David J. Unwin. (2010). Geographic Information Analysis. In *Wiley*. https://doi.org/10.1111/j.1467-9272.2005.00504_11.x

De Gezonde Stad. (2021). Groen. Retrieved February 26, 2021, from https://degezondestad.org/thema/3/groen

Dinther, D. van, & Weijers, E. (2016). *Designing green and blue infrastructure to support healthy urban living*. Retrieved from https://www.researchgate.net/publication/308165682

Doiron, D., Setton, E. M., Shairsingh, K., Brauer, M., Hystad, P., Ross, N. A., & Brook, J. R. (2020). Healthy built environment: Spatial patterns and relationships of multiple exposures and deprivation in Toronto, Montreal and Vancouver. *Environment International*, *143*(April). https://doi.org/10.1016/j.envint.2020.106003

Fan, Y., Das, K. V., & Chen, Q. (2011). Neighborhoud green, social support, physical activity, and stress: Assesing the cumulative impact. *Health and Place*, *17*(6), 1202–1211. https://doi.org/10.1016/j.healthplace.2011.08.008

Gemeente Amsterdam. (2016). *Koers 2025*. Amsterdam.

Gemeente Amsterdam. (2017). *Meerjarenplan fiets 2017-2022*. Amsterdam.

Gemeente Amsterdam. (2020). *Groenvisie 2050: Een Leefbare Stad Voor Mens en Dier*. Amsterdam.

Gemeente Amsterdam. (2021). De stad wordt radicaal groener. Retrieved February 26, 2021, from https://www.amsterdam.nl/nieuws/nieuwsoverzicht/radicaal-groener/

Green City Solutions. (2021). High-tech trees for clean air. Retrieved February 28, 2021, from investEU website: https://europa.eu/investeu/projects/high-tech-trees-clean-air_en

Gunn, L. D., Mavoa, S., Boulangé, C., Hooper, P., Kavanagh, A., & Giles-Corti, B. (2017). Designing healthy communities: Creating evidence on metrics for built environment features associated with walkable neighbourhood activity centres. *International Journal of Behavioral Nutrition and Physical Activity*, *14*(1), 1–12. https://doi.org/10.1186/s12966-017-0621-9

Harbers, A., Spoon, M., Amsterdam, H. van, & Schuit, J. van der. (2019). *Ruimtelijke Dichtheden En Functiemenging in Nederland (RUDIFUN)*.

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

Holtan, M. T., Dieterlen, S. L., & Sullivan, W. C. (2015). Social Life Under Cover: Tree Canopy and Social Capital in Baltimore, Maryland. *Environment and Behavior*, *47*(5), 502–525. https://doi.org/10.1177/0013916513518064

Jennings, V., & Bamkole, O. (2019). The relationship between social cohesion and urban green space: An avenue for health promotion. *International Journal of Environmental Research and Public Health*, *16*(3). https://doi.org/10.3390/ijerph16030452

Kleemans, F. (2014). *Invloed van sociaal kapitaal op vrijwilligerswerk*. Universiteit Utrecht.

Klompmaker, J. O., Hoek, G., Bloemsma, L. D., Gehring, U., Strak, M., Wijga, A. H., … Janssen, N. A. H. (2018). Green space definition affects associations of green space with overweight and physical activity. *Environmental Research*, *160*(October 2017), 531–540. https://doi.org/10.1016/j.envres.2017.10.027

Koops, R. (2021, February 15). GroenLinks: afwaardering A10 West maakt bouw woningen mogelijk. *Het Parool*, p. 2. Retrieved from https://www.parool.nl/amsterdam/groenlinks-afwaardering-a10-west-maakt-bouw-woningen-mogelijk~b1029c74/

KPMG. (2012). *Groen, gezond en productief. the Economics of Ecosystems & Biodiversity (tEEB NL): natuur en gezondheid*. 39. Retrieved from kpmg.nl

Kumar, P., Druckman, A., Gallagher, J., Gatersleben, B., Allison, S., Eisenman, T. S., … Morawska, L. (2019). The nexus between air pollution, green infrastructure and human health. *Environment International*, *133*(June), 105181. https://doi.org/10.1016/j.envint.2019.105181

Li, F., Zhou, T., & Lan, F. (2021). Relationships between urban form and air quality at different spatial scales: A case study from northern China. *Ecological Indicators*, *121*(September 2020), 107029. https://doi.org/10.1016/j.ecolind.2020.107029

Li, T., Fang, Y., Zeng, D., Shi, Z., Sharma, M., Zeng, H., & Zhao, Y. (2020). Developing an indicator system for a healthy City: Taking an urban area as a pilot. *Risk Management and Healthcare Policy*, *13*, 83–92. https://doi.org/10.2147/RMHP.S233483

Lopez, R. (2004). Urban sprawl and risk for being overweight or obese. *American Journal of Public Health*, *94*(9), 1574–1579. https://doi.org/10.2105/AJPH.94.9.1574

Makhatkov, I. (2019). On the validation of models. *BIO Web of Conferences*, *16*, 1–4. https://doi.org/10.1080/02723646.1981.10642213

Milgen, A. van. (2016). Urban Heat Island And Urban Configuration. *Centre for Geo-Information*.

Mouratidis, K. (2020). Neighborhood characteristics, neighborhood satisfaction, and well-being: The links with neighborhood deprivation. *Land Use Policy*, *99*(June), 104886. https://doi.org/10.1016/j.landusepol.2020.104886

Mowafi, M., Khadr, Z., Bennett, G., Hill, A., Kawachi, I., & Subramanian, S. V. (2012). Is access to neighborhood green space associated with BMI among Egyptians? A multilevel study of Cairo neighborhoods. *Health and Place*, *18*(2), 385–390. https://doi.org/10.1016/j.healthplace.2011.12.002

Mueller, N., Rojas-Rueda, D., Khreis, H., Cirach, M., Andrés, D., Ballester, J., … Nieuwenhuijsen, M. (2020). Changing the urban design of cities for health: The superblock model. *Environment International*, *134*(September 2019), 105132. https://doi.org/10.1016/j.envint.2019.105132

Muncipality of Amsterdam. (2021). Valley. Retrieved February 26, 2021, from https://zuidas.nl/project/valley/

Murali, M., Cummings, C., Feyertag, J., Gelb, S., Hart, T., Khan, A., … Lucci, P. (2018). 10 Things To Know About the Impacts of Urbanisation. *OBI*, 10. Retrieved from https://www.odi.org/publications/11218-10-things-know-about-impacts-urbanisation

Nematchoua, M. K., Sevin, M., & Reiter, S. (2020). Towards sustainable neighborhoods in Europe: Mitigating 12 environmental impacts by successively applying 8 scenarios. *Atmosphere*, *11*(6). https://doi.org/10.3390/atmos11060603

Nijs, T. de. (2020). Waarde van Groen & Blauw in Amsterdam. In *De Urbanisten*. Amsterdam.

Nørregaard, A. J., Gram, M., Vigelsoe, A., Wiuff, C., & Birk, A. (2014). Distance to Green Space and Physical Activity: A Danish National Representative Survey. *Journal of Physical Activity & Health*, 1–26.

Norton, B. A., Coutts, A. M., Livesley, S. J., Harris, R. J., Hunter, A. M., & Williams, N. S. G. (2015). Planning for cooler cities: A framework to prioritise green infrastructure to mitigate high temperatures in urban landscapes. *Landscape and Urban Planning*, *134*, 127–138. https://doi.org/10.1016/j.landurbplan.2014.10.018

OIS. (2018). Grote groenonderzoek. *OIS Amsterdam*, *0*.

Orstad, S. L., Szuhany, K., Tamura, K., Thorpe, L. E., & Jay, M. (2020). Park proximity and use for physical activity among urban residents: Associations with mental health. *International Journal of Environmental Research and Public Health*, *17*(13), 1–13. https://doi.org/10.3390/ijerph17134885

Patel, S. B. (2013). Life between buildings: The use and abuse of FSI. *Economic and Political Weekly*, *48*(6), 68–74.

Paulin, M., Remme, R. P., & de Nijs, T. (2019). *Amsterdam's Green Infrastructure: Valuing Nature's Contributions to People*.

Peen, J., Schoevers, R. A., Beekman, A. T., & Dekker, J. (2010). The current status of urban-rural differences in psychiatric disorders. *Acta Psychiatrica Scandinavica*, *121*(2), 84–93. https://doi.org/10.1111/j.1600-0447.2009.01438.x

Plane, J., & Klodawsky, F. (2013). Neighbourhood amenities and health: Examining the significance of a local park. *Social Science and Medicine*, *99*, 1–8. https://doi.org/10.1016/j.socscimed.2013.10.008

Rijksinstituut voor Volksgezondheid en Milieu. (2018). *Integratiematen voor de Volksgezondheid Toekomst Verkenning (VTV) 2018*.

Rijkswaterstaat. (2020). Online bezoekerscentrum Gaasperdamtunnel. Retrieved February 26, 2021, from https://bezoekerscentrum.rijkswaterstaat.nl/SchipholAmsterdamAlmere/news/de-stappen-van-de-openstelling-van-de-gaasperdammertunnel/#.YDgFimhKjHo

Royal FloraHolland. (2021). *Groen en Wonen: De Meerwaarde van Groen in de Stedelijke Omgeving*. https://doi.org/10.1038/srep11610

Silvennoinen, K. C. (2017). *Influence of Urban Design in the Choice of Transportation Mode – Cycling for a People-Centred Urban Form*. Utrecht Univeristy.

Staatsen, B., van der Vliet, N., Kruize, H., Hall, L., Morris, G., Bell, R., & Stegeman, I. (2017). *Exploring triple-win solutions for living, moving and consuming that encourage behavioural change, protect the environment, promote health and health equity*. Retrieved from http://inherit.eu/wp-content/uploads/2017/06/INHERIT-Report-A4-Low-res_s.pdf

Tobler, W. . R. . E. G. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234–240.

Trencher, G., & Karvonen, A. (2019). Stretching “smart”: advancing health and well-being through the smart city agenda. *Local Environment*, *24*(7), 610–627. https://doi.org/10.1080/13549839.2017.1360264

UN Habitat. (2020). *World Cities Report 2020: The Value of Sustainable Urbanization*. https://doi.org/10.18356/c41ab67e-en

Vardoulakis, S., Salmond, J., Krafft, T., & Morawska, L. (2020). Urban environmental health interventions towards the Sustainable Development Goals. *Science of the Total Environment*, *748*, 141530. https://doi.org/10.1016/j.scitotenv.2020.141530

Wesseling, J., Beijk, R., & Kuijeren, N. Van. (2008). Effecten van groen op de luchtkwaliteit. *RIVM*, 1–37.

Westfall, M. S., & Villa, V. a De. (2001). *Urban Indicators for Managing Cities*. Asian Development Bank.

World Health Organisation. (2010). *Urban planning, environment and health: From evidence to policy action - meeting report*. Retrieved from http://www.euro.who.int/__data/assets/pdf_file/0004/114448/E93987.pdf%5Cnpapers2://publication/uuid/D1F4723F-9856-4156-8C4D-5DB0EE873300

World Health Organisation. (2020a). Health effects of urban environments. Retrieved February 26, 2021, from WHO website: https://www.euro.who.int/en/health-topics/environment-and-health/Housing-and-health/activities/health-effects-of-urban-environments#:~:text=Health challenges in the urban,reduced opportunities for physical activity.

World Health Organisation. (2020b). Urban Health. Retrieved February 26, 2021, from https://www.who.int/health-topics/urban-health

Zanen, K. Van, Ponteyn, B., & Keijzer, E. (2011). *Structuurvisie Amsterdam 2040 Economisch sterk en duurzaam*. Retrieved from www.amsterdam.nl/publish/.../structuurvisie_def_maart2011_web.pdf

GIMA
Geographical Information Management and Applications

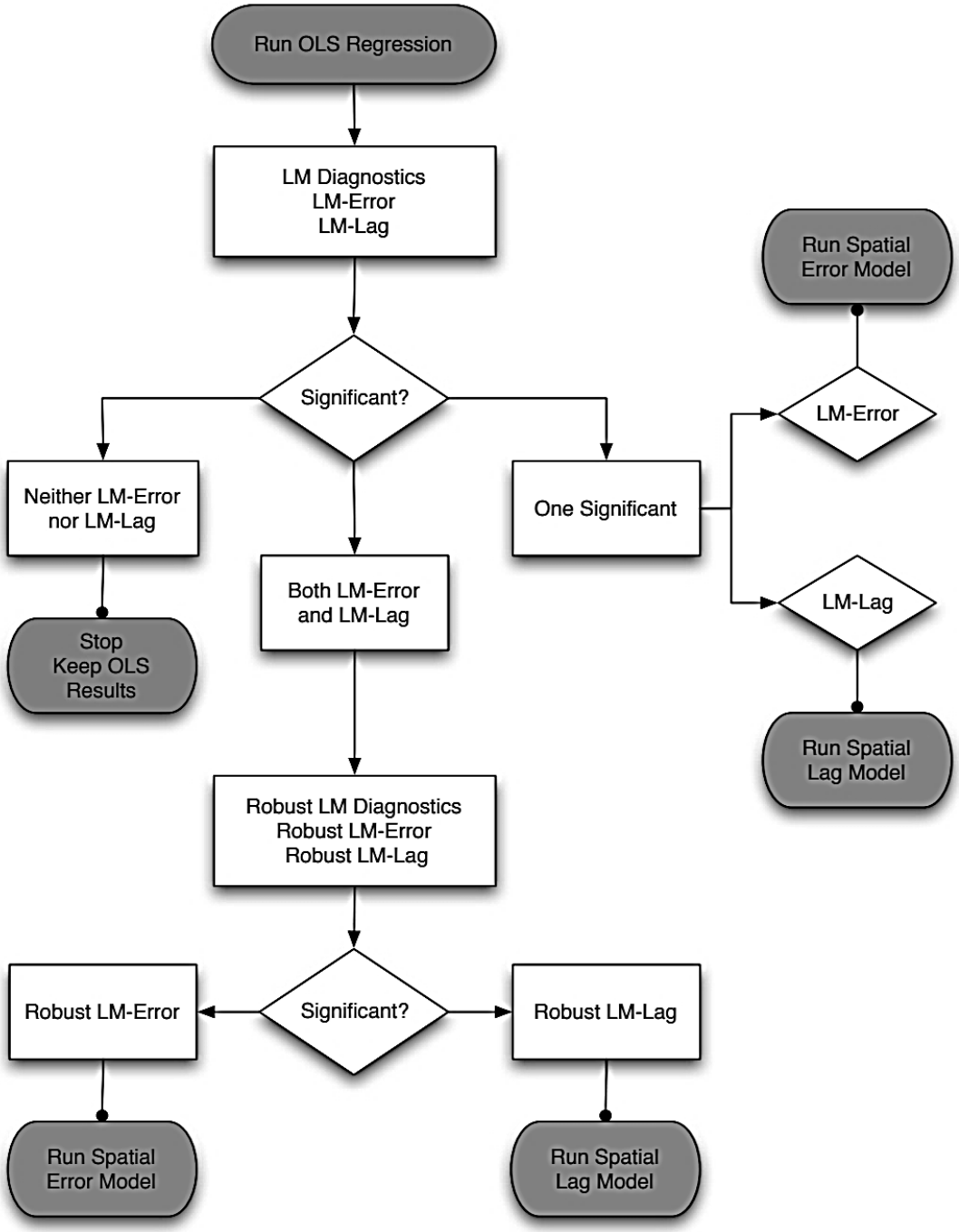
The impact of Urban Green Infrastructure on health: a Amsterdam case-study

Appendices

Universiteit Utrecht
TU Delft
UNIVERSITY OF TWENTE.
WAGENINGEN UR

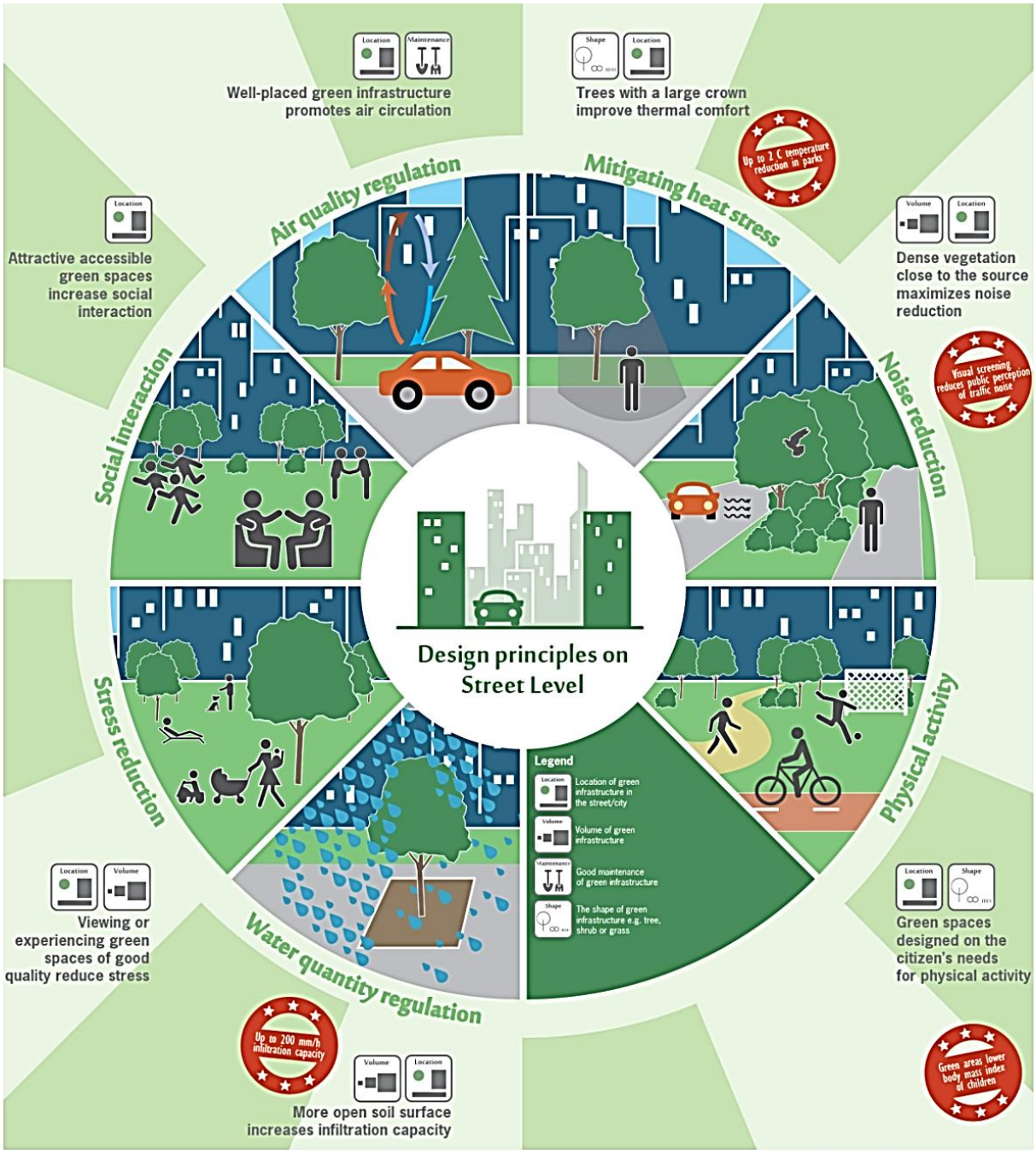
Appendix 1

Regression model decision-tree (Anselin, 2005, p. 199)



Appendix 2

Infographic showing the health benefits of PUGI in urban environments on street level/direct environment of citizens (Bosch & Goossen, 2016, p.16)



Appendix 3

Table showing all datasets, source, general information, and pre-processing steps

Dataset untranslated name (format) (type)	*Source of metadata (weblink)*	*info*	*Preparation steps*
Functiemix (MIF) (Nominal)	https://maps.amsterdam.nl/open_geodata/?k=152	Land use of the municipality of Amsterdam categorized by CBS_2 land use profiles, 2017 data.	-Converting the original file into a shapefile using FME workbench. -Filtering only the public urban green land uses, namely CBS_code2: 40, 60, 61, 52, 70, 75, 78 (CBS, 2008) in ArcGisPro. These CBS_code2 classifications consisted of the more precise CBS_code1 classes. But after inspection the CBS_code2 classes were the most useful for categorization.
Fietsnetwerk Amsterdam (SHP) (Nominal)	https://data.overheid.nl/dataset/196aeef9-eeba-4442-b2df-1ca3e9f9f9f1	Dataset showing all separated and marked bike-lanes in the municipality of Amsterdam. 2016 data.	-The median width of all bicycle paths is exactly 2 meters; therefore a 1 meter buffer will be placed around the bicycle paths to quantify its average surface area. -Also, the polyline segments are dissolved to create 1 large infrastructure.
Groene daken (MIF) (Ratio)	https://maps.amsterdam.nl/open_geodata/?k=51	All registered green roofs and, for most of them their surface area, 2020 data.	-Converting into shapefile using FME workbench. -Filtering out green roofs with no surface area because they are not quantifiable in the regression later (345/457 elements remaining). -To calculate the radius of the buffer around each point based upon the surface area the formula: radius = √surface-area/∏. -The surface area of the radius is then spatially intersected with the neighbourhoods to see how many square meters of green roof there is within each neighbourhood.
Bomen (MIF) (Nominal)	Dataset 1 (70.000 records): https://maps.amsterdam.nl/open_geodata/?k=254 Dataset 2 (70.000 records): https://maps.amsterdam.nl/open_geodata/?k=255 dataset 3 (70.000 records): https://maps.a	Contains 4 datasets with all 259.431 public trees in Amsterdam and their location, type, and size. Private green infrastructure is not considered as they are not freely accessible and thus not part of the *public* green infrastructure. Think: Zoo, private gardens, private parking lot or industrial lot, 2020 data.	-Converting into shapefile using FME workbench. -Combining the 4 datasets with reach around 70.000 entries into one using ArcGisPro. -Then intersect the trees by the 99 neighbourhoods to eliminate trees outside of the neighbourhoods, resulting in 252.771 final entries.

	msterdam.nl/open_geodata/?k=256 dataset 4 (49.431 records): https://maps.amsterdam.nl/open_geodata/?k=257		
Gebied Buurtcombinaties (MIF) (Nominal)	https://maps.amsterdam.nl/open_geodata/?k=200	All 99 neighbourhoods (called *buurtcombinaties/wijken* in Dutch), 2020 data.	-Converting into shapefile using FME workbench. -Within ArcGIS this shapefile will be sued to spatially join all the PUGI with the 99 (91 after filtering) corresponding neighbourhoods for later usage to GeoDa.
Bevolking wijken 2020-2050 (Excel) (Ratio)	https://data.amsterdam.nl/datasets/DMknRs8hEH-CtA/bevolking-wijken/	The estimate of inhabitants per neighbourhood for the following years: 2020, 2025., 2030, 2040 and 2050. (2020 data)	-Convert the excel table into a CVS to make them compatible with ArcGisPro. -Within ArcGisPro Join them based upon the unique wijkcode with the wijken shapefile in ArcGisPro.
Parkeervlakken (SHP) (Nominal)	https://data.amsterdam.nl/datasets/D6rMG5CdGBfp2Q/	Categorized database of all parking spaces in the municipality of Amsterdam; the categories differentiate between taxi parking, regular parking, handicap parking etcetera. This information can later be used to create the 2050 green infrastructure estimate, 2017 data.	-The parkeervlakken will be used in RQ3, RQ4 and RQ5 to see if turning the street-parking into green infrastructure
TOP10NL (SHP) (Nominal)	http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/29d5310f-dd0d-45ba-abad-b4ffc6b8785f	Detailed map of the ground features in the netherlands. For Amsterdam only the data from tiles 25W and 25O are needed.	-this map will be used as a potential background map for the maps being made and as a reference for checking the position of the variables.
Depression and loneliness statistics (Excel) (Ratio)	https://statline.rivm.nl/#/RIVM/nl/dataset/50052NED/table	percentage of people that experience light depression and high depression symptoms + percentage of people with light loneliness and high loneliness symptoms.	-turb the excel file into a CVS for interoperability with ArcGisPro -link the health statistics to the PUGI-map layer based upon the neighbourhood coding,

Obesity and overweight statistics (Excel) (Ratio)	https://statline.rivm.nl/#/RIVM/nl/navigatieScherm/thema	Health variables and on a neighbourhood, district, and city level, 2016 data.	-The file was downloaded in excel format. The Buurtcombinaties coding (T92, M52 etc.) system was added to make them better joinable in ArcGisPro; as sometimes the neighbourhood names contain small typos (Capital letters, connection line (-) etc that are not present in the RIVM data). -import as CSV in ArcGisPro and join with the PUGI dataset
Openbaar Vervoer haltes (SHP) (Nominal)	https://www.arcgis.com/home/item.html?id=68d84ef791594f329dc4b557645090a9	Point data of all public transport stops (bus, taxi, tram, ferry, and train)	-ready to use shapefile that needs to be aggregated to the neighbourhood and neighbourhood buffers.
RUDIFUN: Ruimtelijke dichtheden en Functie-menging Nederland (SHP) (Ratio)	https://data.overheid.nl/dataset/rudifun-ruimtelijke-dichtheden-en-functiemenging-nederland-pbl	Density data on different aggregation levels, from housing block to provincial. Data in net as well as gross FSI.	-ready to use shapefile -the neighbourhood names and codes have a discrepancy with the codes and names used by the municipality of Amsterdam, and also their spatial extent differs. Therefore, the building block net FSI was aggregated to the official neighbourhood aggregations using spatial joins and averaging out the building block FSI per neighbourhood.

Appendix 4

Table showing important literature in the realm of PUGI and health.

Autor	*Case study location*	*Original Title /* Translated title	*Methods*	*Findings*
(Bos, 2020)	Deventer (Netherlands)	*Urban planning interventions in relation to health and well being*	Expert interviews, resident interviews	The effect of green- interventions differ greatly per neighbourhood. The positive impact of green infrastructure is wider then solely the neighbourhood residents. Besides improving social cohesion in general, also the employers, insurance companies and municipality as a whole can benefit from green improvements.
(Bosch & Goossen, 2016)	Utrecht (Netherlands)	Designing green and blue infrastructure to support healthy urban living	Using the POSAD framework to link health to design principles	Green and blue infrastructure produces many ecosystem services; air temperature regulation, air quality improvement, water regulation, noise reduction, mental health improvements, increasing social interaction and physical activity and mitigation of floods and droughts.
(Congdon, 2019)	Australia, Canada, Korea, UK & USA	Obesity and Urban Environments	Literature review	Low density environments in general lead to less health food in someone's home proximity and thus a higher BMI. People living closer to a public open space (POS) in general have better health in all age groups due to more physical activity. However, the access to POS is related to ethnicity and income, thus these are important covariates. Besides the build environment, the social environment in a neighbourhood (measured by social capital) is also explaining variance in obesity.
(Doiron et al., 2020)	Toronto, Montreal & Vancouver (Canada)	Healthy built environment: Spatial patterns and relationships of multiple exposures and deprivation	Spatial analysis of urban variables (walkability, pollution, greenness, and urban deprivation) to show hotspots/cold spots	The least deprived areas were in general more walkable, had lesser pollution and more green-infrastructure. This result suggests at least in Canada, environmental inequality is prevalent. For these variables, in general socio economic status (SES), built environment and public health are closely connected.

(Fan, Das, & Chen, 2011)	Chicago (USA)	Neighbourhood green, social support, physical activity, and stress: Assessing the cumulative impact	OLS Regression model	Neighbourhood green has several distinct roles in reducing stress; parks mitigate stress indirectly by fostering social support. Park/green size has an overall more positive effect then the overall neighbourhood green level. It is important green spaces in urban areas accommodate socialization opportunities.
(Gunn et al., 2017)	Melbourne (Australia)	Designing healthy communities: creating evidence on metrics for built environment features associated with walkable neighbourhood activity centres	Cluster analysis and multilevel logistic regression	Neighbourhood activity centres (clusters of amenities like shops, barbers, and social life) were analysed on their walkability in regard to their surrounding customers. Outcome was that among other things the connectiveness of the street network, the diversity of the NAC and the residential density impacted the % of residents that went to them walking positively when being high.
(Holtan, Dieterlen, & Sullivan, 2015)	Baltimore (Maryland, USA)	Social Life Under Cover: Tree Canopy and Social Capital	Multiple regression model based upon survey data	There is in this study a positive relationship between the neighbourhood tree canopy cover and the social capital of individuals.
(Jennings & Bamkole, 2019)	Global	The Relationship between Social Cohesion and Urban Green Space: An Avenue for Health Promotion	Literature review	Social cohesion Is associated with increased physical and physiological health benefits. The presence of urban green infrastructure can increase social cohesion and thus enhance health and well-being. Especially the social interaction within urban green infrastructure is enhancing and catalysing social cohesion, social capital, and healthy behaviour.
(Kardan et al., 2015)	Policy: Netherlands. Literature: Global	*The added value of green in the urban environment*	Policy document/ literature review	Green infrastructure leads to health benefits, e.g., less ADHD prevalence, less anxiety disorders, less depression, less cardi-vascular diseases, less heath in summer, less stress, more outdoor physical activity, less water nuisance/flooding, less headaches, and better concentration.
(Klompmaker et al., 2018)	Netherlands	Green space definition affects associations of green space with overweight	Cross-sectional study, logistic regression analysis	Nearest green infrastructure is not associated with obesity or physical activity according to the results. The overall NDVI is however correlated with obesity and physical activity. The

		and physical activity		associations where stronger in less-urban areas and with smaller buffers (importance of proximity).
(Kumar et al., 2019)	Global	The nexus between air pollution, green infrastructure, and human health	Literature / policy study	Rapid urbanization turns cities into concrete landscapes. While inclusion of PUGI may support health, socioeconomic support, and environmental improvements. Strategic placement of PUGI is however not related to reduced air-pollution. Thus urban-greening is not a significant measure to mitigate air-pollution.
(Li, Zhou, & Lan, 2021)	Northern China	Relationships between urban form and air quality at different spatial scales: A case study from northern China	Linear regression/ GWR	The urban-form (level of fragmentation, sparseness, greenery etc.) can have an impact on the air quality. This relation is substantiated on different scale-levels using a regression-model and the different urban forms impacted air quality more when using decreases scale-levels. IN general poly-centric scattered and forested urban forms are resulting in the best air-quality.
(Lopez, 2004)	United States	Urban Sprawl and Risk for Being Overweight or Obese	Multilevel census-data analysis	Sprawl increases the risk of being overweight or obese significantly. For every 1-point (1-100 scale) rise in sprawl overweight population increased .2% and obese population with .5%. Concluding sprawl is one of the drivers of increased BMI.
(Mowafi et al., 2012)	Cairo (Egypt)	Is access to neighbourhood green space associated with BMI among Egyptians? A multilevel study of Cairo neighbourhoods	Multilevel analysis	Studies on the PUGI-health relation have been conducted in developed countries generally, this study studies Cairo neighbourhoods. The study found that PUGI is not related to BMI in Cairo, a conclusion that is contradicting the general findings in developed cities and countries.
(Nørregaard, Gram, Vigelsoe, Wiuff, & Birk, 2014)	Denmark	Distance to Green Space and Physical Activity: A Danish National Representative Survey	Cross-sectional survey analysis	The distance to green-spaces in relation to physical activity and BMI is looked into. Living more then 1-kilometre from PUGI does indeed decrease the odds of physical activity. Also, obesity was more prevalent among people living more then 1-kiloemtre from green-spaces.

(Norton et al., 2015)	Melbourne (Australia)	Planning for cooler cities: A framework to prioritise green infrastructure to mitigate high temperatures in urban landscapes Briony	Literature review and PUGI temperature impact framework.	Global warming in combination with urban developments and urbanisation is potentially deadly due to the resulting weather-extremes. The quantification of the cooling benefits of PUGI is elaborated and a framework is made that prioritises types of PUGI based upon their colling abilities.
(Silvennoinen, 2017)	Helsinki (Finland)	influence of Urban Design in the Choice of Transportation Mode – Cycling for a People-Centred Urban Form	*-Master thesis- Mixed-method: statistical analysis and expert-interviews*	This study discusses to what extent people's decision-making in the choice of transportation mode is influenced by urban design factors. The results show that residential location and daily travel distance influenced the inhabitants' purpose and reason to cycle, but were not significant when cycling was not chosen as a transportation alternative *(quoted from abstract on page 5)*
(Vardoulakis, Salmond, Krafft, & Morawska, 2020)	Global	Urban environmental health interventions towards the Sustainable Development Goals Sotiris	Review of 14 case-studies	The SDG's aim to increase sustainable lifestyles by 2030, due to the majority of the people living in urban-environments, implanting urban interventions that will improve health, wellbeing and sustainability in the built environment is essential. Promoting climate-sensitive urban planning and policy is an important means to improve health; think zero-carbon urban housing, active transport, and heat and flood protection.

Appendix 5

Table showing alle relevant datasets and their details and substantiation.

Independent variables (PUGI)	*Unit*	*Standardization*	*Year(s)*	*Substantiation*
Dedicated bicycle paths	m bike lane / m² neighbourhood	yes	Current + 2050 estimate	(Plane & Klodawsky, 2013), (Vardoulakis et al., 2020), (Gemeente Amsterdam, 2017)
*Total public urban **green** (PUGI):* public parks (CBS 40), forest (CBS 60) , public sport facilities (CBS 41), recreational areas (CBS 43) and miscellaneous natural areas (CBS 61 & CBS 62).	m² green / m² neighbourhood	Yes	Current + 2050 estimate	(World Health Organisation, 2020), (Staatsen et al., 2017), (Orstad, Szuhany, Tamura, Thorpe, & Jay, 2020), (Zanen, Ponteyn, & Keijzer, 2011)
*Total public urban **blue** infrastructure: water > 6 meter (CBS 78), recreational water (CBS 75) and IJsselmeer (CBS 70).*	m² blue / m² neighbourhood	Yes	Current + 2050 estimate	(Vardoulakis et al., 2020), (Bos, 2020), (Kardan et al., 2015) & (Gemeente Amsterdam, 2020)
Surface area of parking spaces (	m² parking space / m² neighbourhood	Yes	2050 estimate	(Paulin, Remme, & de Nijs, 2019)
Trees	Trees (integer) (+normalized)	Yes	Current + 2050 estimate	(Paulin et al., 2019), (Gunn et al., 2017). (Kumar et al., 2019) & (Mowafi et al., 2012)
Surface area of green roofs	m² green roof / m² neighbourhood	Yes	Current + 2050 estimate	(Paulin et al., 2019)
Public transport stops	nr. stops / km² neighbourhood	Yes	Current +2050 estimate	(Jennings & Bamkole, 2019; World Health Organisation, 2010)
Fast food-locations	nr. Locations / km²	Yes	2020 +2050 estimate	(Mowafi et al., 2012), (Congdon, 2019) & (Lopez, 2004)
Dependent Variables (health)	Unit	Standardization	Year(s)	Substantiation
Overweight persons (BMI >25)	Percentage (%)	No	Current	(Vardoulakis et al., 2020), (World Health Organisation, 2010)
Obese persons (BMI > 30)	Percentage (%)	No	Current	(Vardoulakis et al., 2020), (World Health Organisation, 2010)
Loneliness	Percentage (%)	No	Current	(Bos, 2020), (Holtan et al., 2015), (Jennings & Bamkole, 2019)

Mental health problems	Percentage (%)	No	Current	(Bos, 2020), (Jennings & Bamkole, 2019)
Respiratory deceases	Percentage (%)	No	Current	(Paulin et al., 2019), (Vardoulakis et al., 2020), (World Health Organisation, 2010)
Cardio-vascular diseases rate	Percentage (%)	No	Current	(Paulin et al., 2019), (Vardoulakis et al., 2020), (World Health Organisation, 2010)
Average life expectancy	Years (floating point)	No	Current	(Staatsen et al., 2017), (World Health Organisation, 2010)
Stress	Percentage (%)	No	Current	(Fan et al., 2011), (Bos, 2020), (Holtan et al., 2015), (Jennings & Bamkole, 2019)
Miscellaneous datasets	Unit	Standardization	Year(s)	Why this parameter?
Current Neighbourhood population and population forecast (2050)	Persons (integer)	No	Current + 2050 estimate	For averaging / general information
Population density	Persons / km²	No	Current + 2050 estimate	"…"
Neighbourhood surface area	m²	No	Current	"…"
Neighbourhood map (99 neighbourhoods)	Neighbourhoods (polygons)		Current	(Nematchoua, Sevin, & Reiter, 2020), (Staatsen et al., 2017)
Address density	Addresses / km²	No	Current	For defining *urban* neighbourhoods
FSI-dataset	Building m²/building lot m²	No	Current	"…"
CBS-vierkanten	Miscellaneous statistics for 100 × 100 and 500 × 500 meter squares	No	Current	Potential aggregation units

Appendix 6

Table showing the abbreviations of all variables in the dataset.

Variable name	*Description*
ID	Neighbourhood ID
CODE	Neighbourhood code
NAME	Neighbourhood name
AREAM2	The surface area of a neighbourhood in m²
ADRSKM2	Address per km²
INW2020	Population projection per neighbourhood for 2020
INW2050	Population projection per neighbourhood for 2050
FSI	Floor space index per neighbourhood
ALC%	Percentage of people that consume alcohol
XTRMALC%	Percentage of people that is classified as an extreme drinker
BMI25	Percentage of people that are overweight, BMI>25
BMI30	Percentage of people that are obese, BMI>30
CIG%	Percentage of people that smokes
LILL%	Percentage of people that suffer from chronic diseases/are ill for a long time (long-illness)
SPRT%	Percentage of people that sports regularly
MOVE%	Percentage of people that moves enough according to the ministry of health metrics
FINPR%	Percentage of people with financial problems
INC1000	Income in 1000's of euros
LINC%	Percentage of low-income residents per neighbourhood
MMHP%	Percentage of people with medium mental health problems
XMHP%	Percentage of people with extreme mental health problems
LO%	Percentage of people that is lonely
VL%	Percentage of people that is very lonely
GRNEIST	Standardized green infrastructure intersecting with the neighbourhood
GR400ST	Standardized green infrastructure intersecting with the neighbourhood + 400 meter buffer
GR800ST	Standardized green infrastructure intersecting with the neighbourhood +800 meter buffer
BLNEIST	Standardized blue infrastructure intersecting with the neighbourhood
BL400ST	Standardized blue infrastructure intersecting with the neighbourhood + 400 meter buffer
BL800ST	Standardized blue infrastructure intersecting with the neighbourhood + 800 meter buffer
BINEIST	Standardized bike infrastructure intersecting with the neighbourhood
BI400ST	Standardized blue infrastructure intersecting with the neighbourhood + 400 meter buffer
BI800ST	Standardized blue infrastructure intersecting with the neighbourhood + 800 meter buffer
PTNEIST	Standardized public transport stops intersecting with the neighbourhood
PT400ST	Standardized public transport stops intersecting with the neighbourhood + 400 meter buffer
PT800ST	Standardized public transport stops intersecting with the neighbourhood + 800 meter buffer

Appendix 7

Table showing all data (geo) portals that are used.

Portal	*Scale*	*URL*	*Funding*	*Input*	*About-URL*
Data.overheid	National	https://data.overheid.nl/alle-data	Ministerie van binnenlands zaken en Koninkrijksrelaties	>180 governmental organisations	https://data.overheid.nl/ondersteuning/algemeen/over-dataoverheidnl
CBS statline	National	https://opendata.cbs.nl/statline/#/CBS/nl/	Ministerie van Economische Zaken en Klimaat	CBS (independent organisation)	https://www.cbs.nl/nl-nl/over-ons/organisatie
Data. Amsterdam	Amsterdam	https://maps.amsterdam.nl/open_geodata/?LANG=nl	Municipality of Amsterdam	Municipal input (OIS) + external partners	https://data.amsterdam.nl/artikelen/artikel/about-ois/6d02564e-c3fe-419b-8f4a-381583acc915/
RIVM statline	National	https://statline.rivm.nl/#/RIVM/nl/	Ministerie van volksgezondheid, Welzijn en Sport	The responsible ministry	
PDOK/Nationaal Georegister	National	https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/home	Ministerie van infrastructuur en Waterstaat, ministerie binnenlandse zaken en koninkrijksrelaties, ministerie van economische zaken en klimaat, Rijkswaterstaat, Geonovum	>450 organisations	https://www.pdok.nl/over-pdok
ArcGIS University of Groningen portal	National	https://www.arcgis.com/home/search.html?q=owner%3A%22universityofgroningen%22&t=content&restrict=false	University of Groningen (RUG)	University of Groningen	https://www.rug.nl/society-business/centre-for-information-technology/research/services/gis/portfolio

Appendix 8

Original Dutch quote from which a translation is used in the thesis.

'Groen is belangrijk voor een fijne en gezonde leefomgeving en is hard nodig om de stad te beschermen tegen droogte, hitte en extreme regenval. Groen is voor Amsterdam van levensbelang. In een groeiende stad waar er steeds meer vraag is naar woonruimte en de openbare ruimte onze achtertuin is, is er daarom aandacht nodig voor aanleg, behoud en beheer van groen, zodat iedereen daarvan kan blijven genieten.' (Wijten, 2020, p.5)

Appendix 9

Regression 1.0: {BMI>30; OLS; no weights; Neighborhood}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :       BMI30  Number of Observations:   91
Mean dependent var  :     12.2088  Number of Variables   :    7
S.D. dependent var  :     3.90974  Degrees of Freedom    :   84

R-squared           :    0.686358  F-statistic           :     30.6369
Adjusted R-squared  :    0.663955  Prob(F-statistic)     :3.22438e-019
Sum squared residual:     436.286  Log likelihood        :    -200.442
Sigma-square        :     5.19388  Akaike info criterion :     414.884
S.E. of regression  :     2.27901  Schwarz criterion     :      432.46
Sigma-square ML     :     4.79435
S.E of regression ML:      2.1896

--------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------
          CONSTANT        10.116       0.980984        10.3121      0.00000
               FSI      -4.59325       0.664701       -6.91025      0.00000
             LINC_      0.501862      0.0564362        8.89256      0.00000
           PTNEIST     -0.724666       0.290463       -2.49487      0.01456
           GRNEIST     -0.429561       0.269973       -1.59113      0.11534
           BLNEIST      0.453253       0.296908        1.52658      0.13062
           BINEIST      0.137277       0.316155       0.434208      0.66525
--------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.589815
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE              PROB
Jarque-Bera             2             3.9793          0.13674

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE              PROB
Breusch-Pagan test      6            27.0301          0.00014
Koenker-Bassett test    6            19.8061          0.00300
```

Appendix 10

Regression 1.1 {BMI >30; OLS; K5 weights; -BINEIST; Neighborhood}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :       BMI30  Number of Observations:   91
Mean dependent var  :     12.2088  Number of Variables   :    6
S.D. dependent var  :     3.90974  Degrees of Freedom    :   85

R-squared           :    0.685654  F-statistic           :     37.0806
Adjusted R-squared  :    0.667163  Prob(F-statistic)     :5.47822e-020
Sum squared residual:     437.265  Log likelihood        :    -200.544
Sigma-square        :      5.1443  Akaike info criterion :     413.088
S.E. of regression  :      2.2681  Schwarz criterion     :     428.153
Sigma-square ML     :     4.80511
S.E of regression ML:     2.19206

--------------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------------
          CONSTANT       10.1945       0.959548        10.6243     0.00000
           PTNEIST     -0.654161       0.239686       -2.72924     0.00771
           GRNEIST     -0.379809       0.243282       -1.56119     0.12220
           BLNEIST      0.428214       0.289861        1.47731     0.14329
               FSI      -4.61872       0.658939       -7.00933     0.00000
             LINC_      0.497653      0.0553312        8.99407     0.00000
--------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.414575
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE              PROB
Jarque-Bera             2             4.9408          0.08455

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE              PROB
Breusch-Pagan test      5            22.6123          0.00040
Koenker-Bassett test    5            16.2072          0.00628

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                          MI/DF        VALUE          PROB
Moran's I (error)             0.2566        4.9438        0.00000
Lagrange Multiplier (lag)       1          27.2619        0.00000
Robust LM (lag)                 1          10.6738        0.00109
Lagrange Multiplier (error)     1          16.9584        0.00004
Robust LM (error)               1           0.3703        0.54284
Lagrange Multiplier (SARMA)     2          27.6322        0.00000
```

Appendix 11

Regression 1.2 {BMI >30; lag-model; Q1 weights; -BINEIST; Neighborhood}

```
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Geoda_export_11022021
Spatial Weight      : Geoda_export_11022021_queen1st
Dependent Variable  :       BMI30  Number of Observations:   91
Mean dependent var  :     12.2088  Number of Variables   :    7
S.D. dependent var  :     3.90974  Degrees of Freedom    :   84
Lag coeff.   (Rho)  :    0.486869

R-squared           :    0.789065  Log likelihood        :    -185.158
Sq. Correlation     : -            Akaike info criterion :     384.316
Sigma-square        :     3.22437  Schwarz criterion     :     401.892
S.E of regression   :     1.79565

------------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
           W_BMI30      0.486869      0.0788492        6.17469     0.00000
          CONSTANT       3.38542        1.27019        2.66527     0.00769
           PTNEIST     -0.609779       0.190844       -3.19518     0.00140
           GRNEIST     -0.252123       0.193128       -1.30547     0.19173
           BLNEIST      0.172859       0.235551        0.73385     0.46304
               FSI      -2.74551        0.58412       -4.70025     0.00000
             LINC_      0.433223       0.046151        9.38709     0.00000
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                    DF      VALUE        PROB
Breusch-Pagan test                      5        24.6292     0.00016

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Geoda_export_11022021_queen1st
TEST                                    DF      VALUE        PROB
Likelihood Ratio Test                   1        30.7720     0.00000
```

Appendix 12

Regression 2.0 {BMI >30; OLS; K5 weights; Neighbourhood+400}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :       BMI30  Number of Observations:   91
Mean dependent var  :     12.2088  Number of Variables   :    7
S.D. dependent var  :     3.90974  Degrees of Freedom    :   84

R-squared           :    0.651018  F-statistic           :     26.1166
Adjusted R-squared  :    0.626090  Prob(F-statistic)     :2.58015e-017
Sum squared residual:     485.446  Log likelihood        :      -205.3
Sigma-square        :     5.77912  Akaike info criterion :       424.6
S.E. of regression  :     2.40398  Schwarz criterion     :     442.176
Sigma-square ML     :     5.33457
S.E of regression ML:     2.30967

--------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------
          CONSTANT        10.559       0.971104        10.8732     0.00000
               FSI      -5.06392       0.608082        -8.3277     0.00000
             LINC_      0.502292      0.0594097        8.45471     0.00000
           BI400ST     -0.172625        0.42422      -0.406922     0.68510
           PT400ST     -0.004436       0.441233     -0.0100536     0.99204
           GR400ST      0.206471       0.296243       0.696965     0.48775
           BL400ST      0.260128       0.301188       0.863675     0.39023
--------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.106766
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE              PROB
Jarque-Bera             2            19.8374           0.00005

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE              PROB
Breusch-Pagan test      6            25.2221           0.00031
Koenker-Bassett test    6            13.9814           0.02984

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.1835        3.6029        0.00031
Lagrange Multiplier (lag)        1          24.1879        0.00000
Robust LM (lag)                  1          16.3517        0.00005
Lagrange Multiplier (error)      1           8.6773        0.00322
Robust LM (error)                1           0.8411        0.35909
Lagrange Multiplier (SARMA)      2          25.0290        0.00000
```

Appendix 13

Regression 2.1 {BMI >30; lag-model; Q1 weights; Neighborhood+400}

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Geoda_export_11022021
Spatial Weight      : Geoda_export_11022021_queenlst
Dependent Variable  :       BMI30  Number of Observations:   91
Mean dependent var  :     12.2088  Number of Variables   :    8
S.D. dependent var  :     3.90974  Degrees of Freedom    :   83
Lag coeff.   (Rho)  :    0.503148

R-squared           :    0.762863  Log likelihood        :      -190.7
Sq. Correlation     : -            Akaike info criterion :     397.401
Sigma-square        :     3.62489  Schwarz criterion     :     417.488
S.E of regression   :     1.90392

------------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
           W_BMI30      0.503148      0.0824173        6.10489      0.00000
          CONSTANT       3.27078        1.35699        2.41032      0.01594
           GR400ST   -0.00536895       0.236253     -0.0227254      0.98187
           BL400ST   -0.00402213       0.239369     -0.0168031      0.98659
           PT400ST     -0.143576       0.350719      -0.409377      0.68226
           BI400ST    -0.0266684       0.337158     -0.0790977      0.93695
             LINC_       0.43742      0.0491684        8.89636      0.00000
               FSI      -2.88736       0.578542       -4.99075      0.00000
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                      DF      VALUE        PROB
Breusch-Pagan test                        6        30.4807     0.00003

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Geoda_export_11022021_queenlst
TEST                                      DF      VALUE        PROB
Likelihood Ratio Test                     1        29.1988     0.00000
================================ END OF REPORT =====================================
```

Appendix 14

Regression 3.0 {BMI >30; OLS; K5 weights; Neighbourhood+800}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :       BMI30  Number of Observations:   91
Mean dependent var  :     12.2088  Number of Variables   :    7
S.D. dependent var  :     3.90974  Degrees of Freedom    :   84

R-squared           :    0.652878  F-statistic           :     26.3316
Adjusted R-squared  :    0.628083  Prob(F-statistic)     :2.07283e-017
Sum squared residual:     482.859  Log likelihood        :    -205.057
Sigma-square        :     5.74832  Akaike info criterion :     424.113
S.E. of regression  :     2.39757  Schwarz criterion     :      441.69
Sigma-square ML     :     5.30614
S.E of regression ML:     2.30351

--------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------
          CONSTANT       10.7811       0.961437        11.2136      0.00000
               FSI      -5.28862       0.600712       -8.80392      0.00000
             LINC_      0.501667      0.0592837        8.46213      0.00000
           PT800ST      0.482258       0.541477       0.890634      0.37567
           BI800ST     -0.474621       0.513329      -0.924594      0.35783
           GR800ST      0.465368       0.308344        1.50925      0.13499
           BL800ST     0.0991873       0.296889       0.334089      0.73915
--------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.066985
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE              PROB
Jarque-Bera             2            21.2721           0.00002

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE              PROB
Breusch-Pagan test      6            22.2396           0.00110
Koenker-Bassett test    6            12.2695           0.05622

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.2026        3.8804        0.00010
Lagrange Multiplier (lag)        1          23.9006        0.00000
Robust LM (lag)                  1          13.4878        0.00024
Lagrange Multiplier (error)      1          10.5762        0.00115
Robust LM (error)                1           0.1633        0.68611
Lagrange Multiplier (SARMA)      2          24.0640        0.00001
================================ END OF REPORT ==================================
```

Appendix 15

Regression 3.1 {BMI >30; lag-model; Q1 weights; Neighbourhood+800}

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Geoda_export_11022021
Spatial Weight      : Geoda_export_11022021_queenlst
Dependent Variable  :       BMI30  Number of Observations:   91
Mean dependent var  :     12.2088  Number of Variables   :    8
S.D. dependent var  :     3.90974  Degrees of Freedom    :   83
Lag coeff.   (Rho)  :     0.49498

R-squared           :    0.763369  Log likelihood        :    -190.494
Sq. Correlation     : -            Akaike info criterion :     396.988
Sigma-square        :     3.61716  Schwarz criterion     :     417.075
S.E of regression   :     1.90188

------------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
          W_BMI30       0.49498      0.0832758        5.94387      0.00000
         CONSTANT       3.46012         1.3809         2.5057      0.01222
            LINC_      0.437158      0.0492464        8.87696      0.00000
              FSI      -2.97946       0.588862        -5.0597      0.00000
          PT800ST       0.30153       0.431634        0.69858      0.48481
          GR800ST      0.227954       0.247203       0.922136      0.35646
          BL800ST     -0.048194        0.23584       -0.20435      0.83808
          BI800ST      -0.26504       0.409421      -0.647353      0.51740
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                    DF      VALUE        PROB
Breusch-Pagan test                      6        29.2323     0.00005

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Geoda_export_11022021_queenlst
TEST                                    DF      VALUE        PROB
Likelihood Ratio Test                   1        29.1256     0.00000
================================ END OF REPORT =====================================
```

Appendix 16

Regression 4.0: {Mental Health Problems; OLS; K5 weights; Neighborhood}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :       MMHP_  Number of Observations:   91
Mean dependent var  :     48.8352  Number of Variables   :    7
S.D. dependent var  :     4.87484  Degrees of Freedom    :   84

R-squared           :    0.752762  F-statistic           :     42.6256
Adjusted R-squared  :    0.735102  Prob(F-statistic)     :1.76471e-023
Sum squared residual:     534.659  Log likelihood        :    -209.693
Sigma-square        :     6.36499  Akaike info criterion :     433.387
S.E. of regression  :     2.52289  Schwarz criterion     :     450.963
Sigma-square ML     :     5.87537
S.E of regression ML:     2.42392

--------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------
          CONSTANT        39.747        1.08596        36.6007      0.00000
               FSI       -3.2319       0.735832       -4.39217      0.00003
             LINC_      0.943945      0.0624756         15.109      0.00000
           BINEIST     -0.316281       0.349987      -0.903694      0.36874
           BLNEIST     -0.837377       0.328681       -2.54769      0.01267
           GRNEIST       0.44666       0.298864        1.49453      0.13879
           PTNEIST   -0.00851288       0.321546     -0.0264749      0.97894
--------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.589815
TEST ON NORMALITY OF ERRORS
TEST                   DF          VALUE             PROB
Jarque-Bera             2          102.9892          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF          VALUE             PROB
Breusch-Pagan test      6           20.9257          0.00189
Koenker-Bassett test    6            6.3373          0.38649

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.0207        0.8243        0.40979
Lagrange Multiplier (lag)        1           1.8117        0.17831
Robust LM (lag)                  1           2.0285        0.15438
Lagrange Multiplier (error)      1           0.1107        0.73938
Robust LM (error)                1           0.3274        0.56717
Lagrange Multiplier (SARMA)      2           2.1391        0.34316
================================ END OF REPORT =====================================
```

Appendix 17

4.1: {Mental Health Problems; OLS; K5 weights; -BINEIST; -PTNEIST; Neighborhood}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :       MMHP_  Number of Observations:   91
Mean dependent var  :     48.8352  Number of Variables   :    5
S.D. dependent var  :     4.87484  Degrees of Freedom    :   86

R-squared           :    0.749148  F-statistic           :      64.208
Adjusted R-squared  :    0.737481  Prob(F-statistic)     :4.96876e-025
Sum squared residual:     542.474  Log likelihood        :    -210.354
Sigma-square        :     6.30784  Akaike info criterion :     430.707
S.E. of regression  :     2.51154  Schwarz criterion     :     443.262
Sigma-square ML     :     5.96125
S.E of regression ML:     2.44157

--------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------
          CONSTANT       39.5121        1.05923        37.3025      0.00000
             LINC_      0.954244      0.0612629        15.5762      0.00000
               FSI      -3.12472       0.725769        -4.3054      0.00004
           BLNEIST     -0.774105       0.320854       -2.41264      0.01796
           GRNEIST      0.339282       0.269159        1.26053      0.21089
--------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.387501
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE              PROB
Jarque-Bera             2            93.6557           0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE              PROB
Breusch-Pagan test      4            20.0183           0.00050
Koenker-Bassett test    4             6.2896           0.17854

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.0165        0.7701        0.44121
Lagrange Multiplier (lag)        1           1.8450        0.17437
Robust LM (lag)                  1           2.2353        0.13489
Lagrange Multiplier (error)      1           0.0698        0.79162
Robust LM (error)                1           0.4601        0.49756
Lagrange Multiplier (SARMA)      2           2.3051        0.31582
```

Appendix 18

Regression 5.0: {Mental Health Problems; OLS; K5 weights; Neighbourhood+400}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :       MMHP_  Number of Observations:   91
Mean dependent var  :     48.8352  Number of Variables   :    7
S.D. dependent var  :     4.87484  Degrees of Freedom    :   84

R-squared           :    0.751435  F-statistic           :     42.3233
Adjusted R-squared  :    0.733680  Prob(F-statistic)     :2.20208e-023
Sum squared residual:     537.529  Log likelihood        :    -209.937
Sigma-square        :     6.39915  Akaike info criterion :     433.874
S.E. of regression  :     2.52965  Schwarz criterion     :      451.45
Sigma-square ML     :     5.90691
S.E of regression ML:     2.43041

--------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------
          CONSTANT       38.6039        1.02187        37.7776      0.00000
               FSI       -2.3058       0.639871       -3.60354      0.00053
             LINC_      0.964168      0.0625155        15.4229      0.00000
           PT400ST      0.727618         0.4643        1.56713      0.12084
           BL400ST     -0.486715       0.316933        -1.5357      0.12837
           GR400ST      0.745366        0.31173        2.39106      0.01903
           BI400ST     -0.346661       0.446398       -0.776574     0.43959
--------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.106766
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE              PROB
Jarque-Bera             2            61.9930           0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE              PROB
Breusch-Pagan test      6            19.0159           0.00414
Koenker-Bassett test    6             6.9230           0.32803

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.0815        1.8336        0.06671
Lagrange Multiplier (lag)        1           2.8236        0.09289
Robust LM (lag)                  1           1.2718        0.25944
Lagrange Multiplier (error)      1           1.7115        0.19080
Robust LM (error)                1           0.1596        0.68954
Lagrange Multiplier (SARMA)      2           2.9832        0.22501
================================ END OF REPORT =================================
```

Appendix 19

Regression 6.0: {Mental Health Problems; OLS; K5 weights; Neighbourhood+800}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :       MMHP_  Number of Observations:   91
Mean dependent var  :     48.8352  Number of Variables   :    7
S.D. dependent var  :     4.87484  Degrees of Freedom    :   84

R-squared           :    0.739018  F-statistic           :     39.6436
Adjusted R-squared  :    0.720377  Prob(F-statistic)     :1.65185e-022
Sum squared residual:     564.381  Log likelihood        :    -212.155
Sigma-square        :     6.71882  Akaike info criterion :      438.31
S.E. of regression  :     2.59207  Schwarz criterion     :     455.886
Sigma-square ML     :     6.20198
S.E of regression ML:     2.49038

--------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------
          CONSTANT        38.613        1.03943        37.1481     0.00000
               FSI      -2.24581       0.649445       -3.45804     0.00086
             LINC_      0.959025      0.0640931         14.963     0.00000
           PT800ST      0.462039       0.585405       0.789264     0.43218
           GR800ST      0.565003       0.333359        1.69488     0.09380
           BL800ST     -0.262656       0.320974      -0.818307     0.41550
           BI800ST     -0.206318       0.554973      -0.371763     0.71100
--------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.066985
TEST ON NORMALITY OF ERRORS
TEST                    DF           VALUE              PROB
Jarque-Bera              2            75.1667           0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                    DF           VALUE              PROB
Breusch-Pagan test       6            22.5752           0.00095
Koenker-Bassett test     6             7.6332           0.26622

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.0982        2.0846        0.03711
Lagrange Multiplier (lag)        1           2.9630        0.08519
Robust LM (lag)                  1           0.9352        0.33352
Lagrange Multiplier (error)      1           2.4860        0.11486
Robust LM (error)                1           0.4582        0.49849
Lagrange Multiplier (SARMA)      2           3.4212        0.18076
================================ END OF REPORT =====================================
```

Appendix 20

Regression 7.0: {Loneliness; OLS; K5 weights; Neighborhood}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :         LO_  Number of Observations:   91
Mean dependent var  :     47.8901  Number of Variables   :    7
S.D. dependent var  :     5.15358  Degrees of Freedom    :   84

R-squared           :    0.658820  F-statistic           :     27.0341
Adjusted R-squared  :    0.634450  Prob(F-statistic)     :1.02138e-017
Sum squared residual:     824.598  Log likelihood        :    -229.407
Sigma-square        :     9.81665  Akaike info criterion :     472.814
S.E. of regression  :     3.13315  Schwarz criterion     :      490.39
Sigma-square ML     :     9.06152
S.E of regression ML:     3.01024

--------------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------------
          CONSTANT       39.9443        1.34864        29.6181      0.00000
               FSI      -3.60421       0.913822        -3.9441      0.00017
             LINC_      0.882829      0.0775878        11.3784      0.00000
           BINEIST     -0.158586       0.434645       -0.364864      0.71613
           PTNEIST     -0.594676       0.399324        -1.4892      0.14018
           GRNEIST     -0.128438       0.371156        -0.34605      0.73017
           BLNEIST     0.0392382       0.408185       0.0961285      0.92366
--------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.589815
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE              PROB
Jarque-Bera             2            71.6058           0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE              PROB
Breusch-Pagan test      6            22.9174           0.00082
Koenker-Bassett test    6             7.5504           0.27293

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.2579        4.9456        0.00000
Lagrange Multiplier (lag)        1          10.4020        0.00126
Robust LM (lag)                  1           0.2979        0.58520
Lagrange Multiplier (error)      1          17.1316        0.00003
Robust LM (error)                1           7.0275        0.00803
Lagrange Multiplier (SARMA)      2          17.4295        0.00016
================================ END OF REPORT ================================
```

Appendix 21

Regression 7.1: {Loneliness; spatial-error model; Q1 weights; Neighborhood}

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Geoda_export_11022021
Spatial Weight      : Geoda_export_11022021_queenlst
Dependent Variable  :          LO_  Number of Observations:   91
Mean dependent var  :    47.890110  Number of Variables   :    7
S.D. dependent var  :     5.153577  Degrees of Freedom    :   84
Lag coeff. (Lambda) :     0.498299

R-squared           :     0.724282  R-squared (BUSE)      : -
Sq. Correlation     : -             Log likelihood        : -222.629931
Sigma-square        :      7.32289  Akaike info criterion :      459.26
S.E of regression   :      2.70608  Schwarz criterion     :     476.836

------------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
          CONSTANT       38.9861        1.36872        28.4836      0.00000
             LINC_      0.875473      0.0687847        12.7277      0.00000
               FSI      -2.54741        1.04417       -2.43966      0.01470
           BINEIST     -0.110741       0.349182      -0.317144      0.75113
           PTNEIST     -0.359785       0.333035       -1.08032      0.28000
           GRNEIST     -0.229108         0.3171       -0.72251      0.46998
           BLNEIST      -0.22414       0.552686      -0.405546      0.68508
            LAMBDA      0.498299       0.109758        4.53999      0.00001
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                      DF      VALUE        PROB
Breusch-Pagan test                        6        32.1711     0.00002

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Geoda_export_11022021_queenlst
TEST                                      DF      VALUE        PROB
Likelihood Ratio Test                     1        13.5543     0.00023
```

Appendix 22

Regression 8.0: {Loneliness; OLS; K5 weights; Neighbourhood+400}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :         LO_  Number of Observations:   91
Mean dependent var  :     47.8901  Number of Variables   :    7
S.D. dependent var  :     5.15358  Degrees of Freedom    :   84

R-squared           :    0.656533  F-statistic           :     26.7609
Adjusted R-squared  :    0.632000  Prob(F-statistic)     :1.34313e-017
Sum squared residual:     830.125  Log likelihood        :    -229.711
Sigma-square        :     9.88244  Akaike info criterion :     473.422
S.E. of regression  :     3.14363  Schwarz criterion     :     490.998
Sigma-square ML     :     9.12225
S.E of regression ML:     3.02031

--------------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------------
          CONSTANT        39.764        1.26989        31.3129      0.00000
               FSI      -3.37879       0.795176       -4.24911      0.00006
             LINC_      0.880161      0.0776889        11.3293      0.00000
           BI400ST     -0.542907       0.554744      -0.978662      0.33056
           GR400ST      0.506003       0.387391        1.30618      0.19506
           BL400ST      0.319835       0.393857        0.81206      0.41905
           PT400ST      0.706387       0.576992        1.22426      0.22428
--------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.106766
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE              PROB
Jarque-Bera             2            67.9060           0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE              PROB
Breusch-Pagan test      6            21.3319           0.00160
Koenker-Bassett test    6             7.1754           0.30493

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.3016        5.6504        0.00000
Lagrange Multiplier (lag)        1          11.3041        0.00077
Robust LM (lag)                  1           0.0045        0.94668
Lagrange Multiplier (error)      1          23.4330        0.00000
Robust LM (error)                1          12.1334        0.00050
Lagrange Multiplier (SARMA)      2          23.4375        0.00001
================================ END OF REPORT ================================
```

Appendix 23

Regression 8.1: {Loneliness; spatial-error model; Q1 weights; Neighbourhood+400}

```
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Geoda_export_11022021
Spatial Weight      : Geoda_export_11022021_queenlst
Dependent Variable  :          LO_  Number of Observations:   91
Mean dependent var  :    47.890110  Number of Variables   :    7
S.D. dependent var  :     5.153577  Degrees of Freedom    :   84
Lag coeff. (Lambda) :     0.499025

R-squared           :     0.729417  R-squared (BUSE)      : -
Sq. Correlation     : -             Log likelihood        : -221.784273
Sigma-square        :      7.18651  Akaike info criterion :     457.569
S.E of regression   :      2.68077  Schwarz criterion     :     475.145

------------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
          CONSTANT       39.0013         1.3043        29.9019      0.00000
             LINC_      0.867409       0.067934        12.7684      0.00000
               FSI      -2.40779       0.953275       -2.52581      0.01154
            GR400ST      0.417721       0.303135          1.378      0.16820
            BL400ST      0.257194       0.393538       0.653543      0.51341
            PT400ST      0.921107       0.473002        1.94736      0.05149
            BI400ST     -0.603484       0.466636       -1.29326      0.19592
             LAMBDA      0.499025       0.109659         4.5507      0.00001
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                      DF      VALUE        PROB
Breusch-Pagan test                        6        29.7940     0.00004

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Geoda_export_11022021_queenlst
TEST                                      DF      VALUE        PROB
Likelihood Ratio Test                     1        15.8535     0.00007
```

Appendix 24

Regression 9.0: {Loneliness; OLS; K5 weights; Neighbourhood+800}

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Geoda_export_11022021
Dependent Variable  :         LO_  Number of Observations:   91
Mean dependent var  :     47.8901  Number of Variables   :    7
S.D. dependent var  :     5.15358  Degrees of Freedom    :   84

R-squared           :    0.657906  F-statistic           :     26.9245
Adjusted R-squared  :    0.633471  Prob(F-statistic)     :1.13975e-017
Sum squared residual:     826.807  Log likelihood        :    -229.529
Sigma-square        :     9.84294  Akaike info criterion :     473.058
S.E. of regression  :     3.13735  Schwarz criterion     :     490.634
Sigma-square ML     :     9.08579
S.E of regression ML:     3.01426

--------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------
          CONSTANT       40.0076        1.25809        31.8002     0.00000
               FSI      -3.55812       0.786064        -4.5265     0.00002
             LINC_      0.874513       0.077576         11.273     0.00000
           BI800ST     -0.816674        0.67172        -1.2158     0.22747
           PT800ST       1.07512       0.708553        1.51735     0.13293
           GR800ST      0.587268       0.403485        1.45549     0.14926
           BL800ST      0.282809       0.388496       0.727958     0.46866
--------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   8.066985
TEST ON NORMALITY OF ERRORS
TEST                    DF           VALUE              PROB
Jarque-Bera              2            69.7397           0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                    DF           VALUE              PROB
Breusch-Pagan test       6            19.7460           0.00307
Koenker-Bassett test     6             6.6650           0.35294

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Geoda_export_11022021_K5
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.3085        5.7018        0.00000
Lagrange Multiplier (lag)        1          11.0182        0.00090
Robust LM (lag)                  1           0.0060        0.93840
Lagrange Multiplier (error)      1          24.5155        0.00000
Robust LM (error)                1          13.5032        0.00024
Lagrange Multiplier (SARMA)      2          24.5214        0.00000
================================ END OF REPORT ==================================
```

Appendix 25

Regression 9.1: {Loneliness; spatial-error; Q1 weights; Neighbourhood+800}

```
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Geoda_export_11022021
Spatial Weight      : Geoda_export_11022021_queenlst
Dependent Variable  :         LO_  Number of Observations:   91
Mean dependent var  :   47.890110  Number of Variables   :    7
S.D. dependent var  :    5.153577  Degrees of Freedom    :   84
Lag coeff. (Lambda) :    0.503089

R-squared           :    0.732203  R-squared (BUSE)      : -
Sq. Correlation     : -            Log likelihood        : -221.367999
Sigma-square        :      7.1125  Akaike info criterion :     456.736
S.E of regression   :     2.66693  Schwarz criterion     :     474.312

------------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
          CONSTANT       39.1596        1.29671        30.1993     0.00000
             LINC_      0.867584      0.0677243        12.8105     0.00000
               FSI      -2.56425       0.955047       -2.68494     0.00725
            PT800ST       1.23619       0.589799        2.09594     0.03609
            GR800ST      0.541785       0.322616        1.67935     0.09308
            BL800ST      0.183436       0.351639       0.521661     0.60191
            BI800ST     -0.837055       0.571774       -1.46396     0.14320
             LAMBDA      0.503089       0.109104         4.6111     0.00000
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                      DF      VALUE        PROB
Breusch-Pagan test                        6        30.5800     0.00003

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Geoda_export_11022021_queenlst
TEST                                      DF      VALUE        PROB
Likelihood Ratio Test                     1        16.3215     0.00005
```