Equity of Food Access in the Netherlands

Associations of Supermarket outlet exposure with socio-economic and demographic characteristics

Abstract

In the last decade, research has increasingly focussed on the influence of the food environment on the dietary choice. Therefore, understanding whether certain socio-economic or demographic groups are disadvantaged in terms of their food envrionment could improve the general health of the population. Multiple regressions and graphical analyses with various measurements for food access are used to determine a relationship in the Netherlands. Generally, food access improves with higher percentages of elderly people or migrants. However, a lower average neighbourhood income is associated with worse supermarket access.

Lukas Tiemann (S3733300)

lukasnetzclub@googlemail.com

Table of Contents

List of Figures

List of Tables

1. Introduction

Currently, about 49.3 per cent of the Dutch adult population is overweight (Statline, 2018). This figure has drastically increased over the past 35 years from only 32 per cent in 1981 (CBS, 2018). Therefore, innovative policy solutions to tackle this problem are necessary.

Since the 2000s, research has increasingly focused upon the influence of the food environment on dietary choice and health outcomes, underlining the importance of food access and availability for the nutritional quality of the diet (Black et al., 2013). Therefore, securing and improving the food access for all socioeconomic and demographic groups may benefit the general health of the population.

Food and grocery stores may be promoting a healthier lifestyle by offering healthy, fresh, organic and locally produced foods. Potential benefits of a diet based on these kinds of foods include normal weight, lower risk of chronic diseases and the reduction of other risk factors (Kawakami et al., 2010). Especially in comparison to the energy dense food of fast food restaurants, supermarkets offer a number of healthy alternatives.

This study's main goal is to analyse the equity of food access for different socioeconomic and demographic groups. In the context of the present study, two main reasons show the importance of such an analysis:

The first reason is best explained by the concept of *Deprivation amplification*. The health of an individuals is impacted by risk factors of obesity which are then further amplified by exposure to a food retail environment offering too few choices of nutritious food (Ver Ploeg et al., 2009). Risk factors such as low-income or migration background are often associated with limited knowledge about nutrition. Therefore, it is assumed that the food environments of low-income subpopulations require special consideration due to the vulnerability of these individuals (Gittersohn & Sharma, 2009). If these groups have worse food access than the average, the food environment may have reinforcing impacts on their diet, resulting in even worse dietary outcomes.

Secondly, certain groups have particular problems in accessing healthy food. For example, low-income households may not be able to buy a car and, thus, can reach fewer destinations. Another example are elderly people who may have physical restrictions preventing them to travel long distances. Understanding whether these groups are disadvantaged could be used as a basis for future policies.

However, even when the assumption of improved health is omitted, understanding whether socioeconomic variables, such as lower income, are an important indicator for accessibility of supermarket is still of importance. Jones & Simmons (1987) give arguments why the location, number and type of close supermarkets and other retail stores *especially* matter for low-income households:

Due to a lack of money, households are restricted in their choice and can only choose one of the cheapest brands. If there is only one supermarket accessible, the number of affordable brands is limited. Consequently, poor as well as elderly people who are not able to drive or cannot afford a car are at the mercy of the nearest supermarkets: Oftentimes merchants make the most money of expensive items and, thus, more deprived areas may have worse supermarket accessibility since these areas are not as profitable (Jones & Simmons, 1987).

Most research to this day regarding the relationship between food access and neighbourhood characteristics has been conducted in the USA, Canada and Australia, clearly identifying “Food Deserts” in the USA (Beaulac et al., 2009). “Food Deserts” are areas experiencing poor access to healthy and affordable food. However, most of these studies focus on single cities identifying food deserts for certain neighbourhoods. A similar approach has already been conducted in the Netherlands for Amsterdam, concluding that no areas are significantly disadvantaged (Helbich et al., 2017).

No studies so far have investigated the distribution of food outlets and its relations to socio-economic and demographic neighbourhood characteristics for the rest of the Netherlands. In addition, transferring existing results to the Netherlands is difficult for a number of reasons. Firstly, compared to the USA, Canada and Australia, there tends to be less income inequality in the Netherlands (OECD, 2017). Income inequality is seen as one of the main reasons for the existence of food deserts in other parts of the world (Ver Ploeg et al., 2009). Secondly, the bicycle is a far more important transportation mode, possibly resulting in a smaller number of reachable destinations for Dutch people. For example, the average Dutch inhabitant owns more than double as many bicycles as a person living in

Canada or the US (Statistics Netherlands, 2015). A supermarket in close proximity may, therefore, be of much higher importance in the Netherlands. Thirdly, European results concerning the relationship between neighbourhood characteristics and food outlet exposure are not clear-cut as will be further explained in chapter 2.3. Consequently, this study may serve as a basis for future studies in the Netherlands.

Specifically, the goal of the analysis is to determine the state of food outlet exposure for the elderly, migrants and low-income households. The main research question driving the analysis reads:

Do neighbourhoods with higher percentages of migrants, elderly or low-income households differ in terms of their access to food outlets compared to the Dutch average?

After answering this question, further research may determine the extent of the relationship for the various Dutch regions.

Data used in this study regarding characteristics of neighbourhoods is provided by CBS (Toelichting Wijk- en Buurtkaart, 2017). In addition, information about the locations of food outlets in the Netherlands was acquired. With the help of GIS software, several measurements for food outlet exposure were developed and calculated. To determine the relationship between the neighbourhood characteristics two main approaches were applied. In the first step, the relationship is analysed graphically with the help of Cumulative Distribution Functions. Afterwards, multiple regression models are set up to test the robustness of the results. Lastly, the results are compared and conclusions for the distribution of food outlets in the Netherlands are drawn.

2. Research State in Europe

While in the United States the literature about food deserts is extensive and the general existence is more or less undisputed, the literature is not as definite in Europe (Beaulac et al., 2009). In a review, Black et al. (2013) summarized the numerous existing reviews about the effect of the food environment on the population. In the US, low-income and ethnic communities live further from the closest store and have, in some cases, worse access to healthy food. In addition, the authors argued that the literature provides evidence for the existence of a relationship between dietary outcomes and environmental exposure in terms of access to supermarkets. Nevertheless, even in the US, not all studies find a significant positive association.

In Europe, the evidence is conflicted and differs significantly between various studies in the same and in different countries. Since the Netherlands assumedly share a lot more cultural as well as structural similarities with the rest of Europe, the literature review focuses on research conducted in Europe.

2.1 Relationship between Access to Supermarkets and Healthy Food Consumption

As much of the relevance for the present research is based on the assumption that food access has a significant impact on the quality of an individual's diet, the following chapter reviews the evidence for this influence in Europe.

In theory, a model put forward by Glanz et al. (2005) links eating patterns of an individual directly to individual variables (such as sociodemographic factors) and environmental variables. Environmental variables are separated into three different aspects. First, the *Organizational Nutrition Environment* includes the effects of the school or work environment. Second, the *Information Environment* describes effects of advertising or other media platforms on eating behaviour. However, while some research analysed theses aspects (e.g. Callaghan et al., 2015), most researchers focussed on the third aspect of the environmental variables: The *Community or Consumer Nutrition Environment* (Black et al., 2013). *Community Nutrition Environment* describes the impact of the type of supermarkets and their accessibility on the diet of a person. *Consumer Nutrition*

Environment argues for the effect of the availability of healthy options and their price on the diet of a person. Only those studies in Europe which examined the relationship between access to supermarkets and dietary intake focussing on these two environments will be considered here. Furthermore, it should be noted that except for one, all studies have been conducted in the UK. Thus, transferring these results to the Netherlands is difficult.

In a study conducted in the Barnsley area of South Yorkshire (England), the authors examined the distance to the nearest supermarket as well as other possible difficulties with grocery shopping on either fruit or vegetable consumption (Pearson et al., 2005). By conducting a survey, the authors gathered data about fruit and vegetable intake combined with socio-demographic and road-travel distance to the nearest supermarket. However, no significant relationship was found between fruit & vegetable intake and supermarket travel distance. Thus, a more important role of cultural influences impacting an individual's diet is suggested.

Similar results were found by White et al. (2004) in a comprehensive study about food deserts in Newcastle (England). In a regression, the authors were unable to demonstrate a relationship between indicators of healthier eating and various factors of the retail food environment. In contrast, demographic, socioeconomic and behavioural factors were much better predictors for an individual's diet.

Macdonald et al. (2011) found some significant associations between proximity to supermarkets and diet patterns or BMI in Glasgow (Scotland). Still, the authors concluded that the distribution of supermarkets does not have a major influence on diet and weight in the UK.

A possible reason why all of these studies did not find significant relationships may be because most residents in urban settings already have quite good access to food stores (Macdonald et al., 2011). Therefore, Macdonald et al. (2011) suggest an approach where possible food deserts are identified before examining relationships on dietary behaviour.

A study conducted in the Republic of Ireland *did find* a statistically significant role for food availability in influencing the diets of individuals while using a similar methodology as the studies discussed previously (Layte et al., 2011). Individuals who live in an environment with more or closer supermarkets have a significantly better diet with regards to cardiovascular risk. These differences may be explained by the study area: Layte et al. (2011) did not focus on just one urban area, as the previous studies did, but on the whole Republic of Ireland. In doing so the sample size increased substantially and rural areas were incorporated. Nevertheless, the authors highlighted that the effect of the food retail environment, while significant, is still small.

In a study set in Paris (France), the authors obtained data on home addresses, food shopping locations, sociodemographic variables as well as health indicators (Drewnoski et al., 2014). While distance to supermarkets was found unrelated to obesity risk, shopping at lower-cost stores was consistently associated with higher obesity risk.

Lastly, a study conducted in Leeds (England) made use of another approach to assess the impact of the food retail environment on people’s health: The study compared food-consumption patterns using surveys before and after the opening of a new large Tesco food store in an area marked by deprivation and a high amount of residents with relatively low income (Wrigley et al., 2003). The diet of the residents improved after the opening of the new supermarket while only by a very small amount.

In conclusion, many studies in Europe did not find any relationship between the food environment and the diet of an individual. Even if studies did identify a significant relationship the impact is only small. Socioeconomic, demographic and behavioural factors seem to be far more important when trying to understand the diet of an individual.

Going forward, this limited relationship should be kept in mind. Nevertheless, Dutch studies are necessary to comprehend the situation in the Netherlands.

2.2 Studies Identifying Food Deserts in European Countries

There is a lack of a general definition of food deserts (Cummins & Macintyre, 2002). For instance, while some studies argue that urban areas which do not have a store of a certain size can be identified as food deserts, other studies include socio-economic factors and the type of the food store in their analysis by focusing on areas where low-income residents are not able to buy affordable and health food (Walker et al., 2010). For this literature review, studies are considered to be food desert studies if no statistical analysis is used or the statistical analysis is not the main component of the study. Instead food deserts are identified by mapping areas and employing certain thresholds about socio-economic factors and supermarket density.

A study set in Nantes (France) mapped the spatial distribution of supermarkets combined with socio-demographic data (Shaw, 2012). With the socio-demographic data, a profile was created which, in theory, included all people having problems with travelling to remote shops. Six areas in Nantes were identified as food deserts after conducting the analysis. The authors highlighted that these areas do not coincide with the officially recognised deprived areas in Nantes. Thus, just identifying areas by income as being deprived, may miss the important feature of food access.

Another study examined whether food deserts exist in Bratislava (Slovakia) (Krizan et al., 2015). The authors employed a number of different approaches for the accessibility of supermarkets to test the robustness of their results. The potential delimitation of food deserts depends strongly on the selection of the indicators such as quality, variability and price of food. However, most residents did not live in an area which can be classified as a food desert. On the contrary: many areas were even identified as food oases.

Lastly, a study set in Leeds (England), Bradford (England) & Cardiff (Wales), in addition to mapping food deserts, made use of spatial interaction models to predict the flows from residential zones to retail destinations (Clarke et al., 2002). The impact on food deserts when new stores are opened was estimated. The authors identified six problematic areas which may be described as food deserts, according to their methodology. Also, the impact of opening a new large store in these areas may have severe consequences on other existing local stores, only exaggerating the existing problem.

The three introduced studies found partly contradicting results. The existence of food deserts seems to be different between regions, even within Europe. Furthermore, all studies considered only a small area and not whole countries. This makes sense since mapping larger areas will become incomprehensible at some point. As a consequence, this study focuses on statistical measurements instead of a mapping approach when analysing the impact of socio-economic indicators on supermarket accessibility in the whole of the Netherlands.

2.3 Influence of Socioeconomic Factors on Indicators of Accessibility of Supermarkets

Lastly, studies in Europe having a similar focus as this analysis are presented. However, the conducted approaches oftentimes differ significantly between the studies and in comparison with this study.

Several studies in Europe have researched the relationship between area deprivation, as measured by indicators such as income, unemployment rates or educational status, and neighbourhood resources including food stores. While these studies analyse statistical relationships, they still significantly depend on how areas are defined as deprived.

A study set in the entirety of Sweden identified three categories of neighbourhood deprivation by using several socioeconomic indicators and estimating their relationship with the accessibility of supermarkets (Kawakami et al., 2010). For each deprivation status prevalence rates for services and goods were calculated. These can be understood as the probability of an area to offer one of the goods, services and resources examined in the study. For food stores, as for most categories, highly and moderately deprived neighbourhoods had a larger probability of having a food or grocery store. Thus, in Sweden more deprived areas do not suffer from worse food accessibility.

Another nationwide study in England also classified areas by deprivation and estimated the relationship with the neighbourhood food environment (Molaodi et al., 2012). As in Sweden, supermarkets were more common in the most deprived compared to the least deprived areas.

A similar approach was used by Macintyre et al. (2008) in Glasgow (Scotland), however, on a much smaller scale. The author's conclusion is ambiguous: There seems to be no clear pattern of supermarket distribution by neighbourhood deprivation level in Glasgow.

A case study in Plymouth (England) compared two highly deprived areas with two of the least deprived areas (Williamson et al., 2017). More households in the most deprived areas were affected by poor access to food retail provision. In addition, a defined healthy food basket had a lower availability in the more deprived areas.

The only research conducted in the Netherlands about supermarket accessibility and food deserts is the previously mentioned study by Helbich et al. (2017). In addition to the mapping of food deserts, the authors researched whether there is a relationship between supermarket accessibility and property prices or share of native Dutch people in Amsterdam. While the authors were able to find discrepancies in accessibility for both indicators, the relevance for people's daily life is presumably only marginal: The general proximity to the nearest supermarket was always relatively low.

In a Danish study, the authors examined associations of supermarket density with average neighbourhood income (Svatisalee et al., 2010). With a negative binomial analysis, the authors could not find evidence for any spatial patterning of supermarkets by area income.

All in all, studies examining whether a relationship between socioeconomic factors and supermarket accessibility exists offer inconsistent results. It does not seem like supermarket accessibility is negatively associated with socioeconomic indicators of a neighbourhood. On the contrary: two large scale studies in Sweden and England identified better supermarket accessibility in more deprived areas (Kawakami et al, 2010; Molaodi et al., 2012).

3. Hypotheses on Relationships between Socioeconomic & Demographic Variables and Supermarket Accessibility

As discussed previously, the main goal of this analysis is to find differences in supermarket access for minority and low-income groups. To understand the location of supermarkets and possible reasons for differences in food access, it is important to understand the economics behind supermarket locations. By explaining the economical theoretical background, a number of hypotheses on possible relationships between supermarket access and socioeconomic indicators are developed. Moreover, theories for differences in food access for elderly and migrants are presented.

3.1 Economic Background of Supermarket Location

Bitler & Haider (2011) developed an economic framework for the existence of food deserts and, in general, food access. This chapter draws heavily from their model and tries to draw conclusions for the analysis at hand. Only additional papers are indicated by a reference since most of the analysis was done by Bitler & Haider (2011). If no reference is given, the analysis refers to Bitler & Haider (2011). The economic analysis consists of four main components: defining relevant food products, the consumer side (demand side), the food retailer side (supply side) as well as the interactions between these factors (market).

The *definition of relevant food products* relates to product availability and how to define the product. In the case of food access, healthy and nutritious food is oftentimes the primary concern. When analysing food accessibility, it must be defined what products are included as healthy and nutritious food and where and how an individual can get these products. For example, individuals might not only shop close to home but also close to work. Furthermore, supermarkets might not be the only option for buying healthy and nutritious food: farmers' markets and speciality shops offer healthy food, too. For this analysis, a detailed definition of food products is not possible: there is no data on the food available at the different outlets as well as no information about where individuals work and could grocery shop. In addition, only supermarkets are included as a source for healthy and nutritious food. While this definition may lack in some parts, it is widely adopted in the literature (e.g. Helbich et al., 2017; Shaw, 2012).

The determinants for the *demand* for healthy food are mostly prices, income and preferences. In theory, the demand for healthy food should decrease if its price increases and increase if the price for the substitute, unhealthy foods, increases. In addition, under the assumption that healthy food is a normal good, the demand should increase with higher income levels. Wealthier people are able to buy higher quantities and higher priced products. Therefore, high-income areas should have more healthy food stores when compared to low-income areas, although, the preferences of groups may alter these results.

The basic determinant for the *supply* side are the costs of running a supermarket such as labour, land, equipment, etc. Supply should decrease, as each one of these costs increases. In theory, labour and land costs have positive effects for low-income groups: As the land and labour prices are lower in low-income neighbourhoods, it may be cheaper to open up supermarkets in these areas. Thus, this effect runs in the opposite direction in comparison to the demand effect.

Another aspect of supermarket *supply* is the existence of fixed costs: A firm has to charge higher prices to be profitable if it experiences higher fixed costs. These fixed costs may differ immensely by the type of area (Ver Ploeg et al., 2009). In inner cities, land prices may be higher and have a greater impact on the total costs compared to supermarkets in rural areas. For some parts of the analysis, only chain-supermarkets are used as indicators. High-volume supermarket chains are expected to charge lower prices compared to single establishments since these can spread fixed costs over a larger number of establishments. However, whether this is the reality in the Netherlands has still to be established.

In Bitler & Haider's (2011) theory the *market* depicts interactions among suppliers and demanders which determine the product availability and price. Oftentimes it is assumed that individuals are price takers which means that they have little effect on quantities, prices or the variety of products. The same is assumed for firms which have no market power, resulting in perfect competition.

All in all, from the determinants of the demand and supply side, the first hypothesis can be developed which will be tested later on.

Hypothesis 1: Due to the lower demand in low-income areas the access to supermarkets is worse compared to high-income areas. However, lower house prices result in better access to supermarkets because of the smaller costs of opening up and operating a store.

3.2 Access for Minority Groups

In general, economic theory suggests that, as long as markets are competitive, a retail firm which does not discriminate should have the same incentive to locate in an area independently from the percentage of minority groups (Ver Ploeg et al., 2009). Furthermore, if there are discriminatory firms operating, the market could even reward non-discriminatory firms which locate in otherwise underserved areas. However, if firms lack good information on food demand in areas with high ethnic minorities, firms might decide against locating in these areas.

Another explanation possible why minorities may have worse access to supermarkets are housing market restrictions limiting minorities ability to move to areas that have better access to supermarkets (Ver Ploeg et al., 2009).

Lastly, in some cases the local government may decide which areas can be used for the development of a new supermarket. If minority groups are underrepresented in the local government, fewer stores might be opened up in areas with high minority percentages.

Hypothesis 2: Due to lack of information on food demand, housing market restrictions and underrepresentation in the local government, migrants have worse access to food stores compared to natives.

3.3 Access for the Elderly

Again, from an economic viewpoint, there is no reason to assume that elderly people have worse supermarket access than younger people. Similar reasons mentioned in the previous chapter do not necessarily apply to elderly people. Nonetheless, it is important to examine whether supermarket access differs for the elderly, since these groups may have trouble to access stores further away. According to a study set in Japan, proximity to supermarkets influences shopping difficulty significantly for elderly people, however, not as much as physical activity restrictions such as not being a car owner or poor eyesight do (Ishikawa et al., 2016).

To set up a hypothesis, it is assumed that the elderly locate themselves closer to supermarkets so they are still able to access food easily. For example, retirement homes may specifically open up close to supermarkets, to give their residents the possibility of arranging their everyday life as independently as possible. However, if this truly is the underlying effect contributing to the access of elderly people cannot be said for certain.

Hypothesis 3: Since elderly people have trouble travelling long distances, they try to locate themselves as closely as possible to the next food outlet.

3.4 Market Power and Spatial Monopolies

One additional aspect of the previously described market is the concept of *market power* (McCann, p. 25, 2013). In a setting where only a few firms are serving a local market, the firms are assumed to have market power and are able to increase the price or restrict the quantity with respect to a situation in perfect competition. Several factors may lead to such market power, including economies of scale, cartels or patents (Das, pp. 293-295, 2007). In the case of supermarkets, geography and space can also confer monopoly power on firms: If the transport cost to travel to a cheaper competitor which is further away than the local supermarket are higher than the price savings, the local supermarket maintains some market power (McCann, p. 24, 2013).

In the analysis, the variety of supermarkets is examined to find out whether such market power exists in the Netherlands in relation to neighbourhood characteristics. In an area where firms have high enough market power, access to food might be restricted due to higher prices or lower available quantities of food. A previous study by Stelder (2012) examined the existence of spatial monopolies for supermarkets in the Netherlands. However, the study did not analyse whether certain socioeconomic groups are more exposed to spatial monopolies than others. In theory, low-income groups might be easier to lock into a spatial monopoly if these groups do not own a car. Also, minority groups and especially recent immigrants, may not know the price differences between supermarkets and just buy at the store closest to them. Supermarkets could take advantage of these restrictions and try to lock these certain groups into spatial monopolies.

Hypothesis 4: Firms use spatial monopolies to lock in minority groups, elderly or low-income group.

4. Data and Methodology

The data used to analyse the relationship between socioeconomic & demographic factors and food accessibility were gathered from two different sources:

First, spatial data on the neighbourhood level was obtained from Statistics Netherlands (CBS) containing information about each neighbourhood in the Netherlands (Toelichting Wijk- en Buurtkaart, 2017). The neighbourhoods are mostly homogeneous in terms of their function (residential, industrial or recreational area) but vary in size and population. Each area is defined by the responsible municipality themselves with Statistics Netherlands coordinating this format nationally. As the latest available information about neighbourhood income was from 2015 and neighbourhood boundaries partially changed in the meantime, only data from 2015 was used (Toelichting Wijk- en Buurtkaart, 2017). The dataset includes demographic variables such as the percentage of people of 65 and over and socioeconomic variables such as the average income per inhabitant. Moreover, the dataset already contains information about the average number of large supermarkets in the proximity of a neighbourhood and the distance to the next large supermarket. A large supermarket is defined as a store with several kinds of daily items and a minimum floor space of 150 square meters.

Second, information about the locations of supermarkets in the Netherlands was prepared and used. The dataset contains information about the store brand and the floorspace & number of employees of each location. Similarly to Helbich et al. (2017), supermarkets were defined as a standard grocery chain which is at least operating 15 stores in the Netherlands. Chains have, on average, higher price competitiveness and a larger variety of products selection compared to single establishments (Mantovani et al., 1997). While there is a variable for floor space available in the sample, after investigations into specific stores the variable showed a high error rate. Consequently, no floorspace threshold was chosen and instead the analysis focuses only on chain stores. In addition, all "to go" convenience stores which can be found at locations such as airports were removed from the sample, as these stores often only provide ready-to-eat food at non-competitive prices (Helbich et al., 2017). A similar reason applies to organic supermarkets: While these stores do offer healthy food, the pricing in most cases is not competitive and, thus, organic supermarkets were excluded from the analysis (Zenk, et al., 2005). Other smaller adjustments can be found in the appendix (see Appendix 1 – Adjustments to the Supermarket Dataset).

4.1 Descriptive Statistics

Table 1 highlights the indicators used from the Statistic Netherlands "Wijk- en Buurtkaart", showing the sample size, mean, standard division, median, minimum and maximum of each variable. Hereafter, variables will be highlighted where the interpretation may be difficult. For additional information see the explanation of the "Toelichting Wijk- en Buurtkaart 2015, 2016 en 2017" provided by CBS.

The information about migration is divided into the *percentage of people with a western migration* which includes people originally from Europe (except for Turkey), North-America, Ozeania, Indonesia and Japan and the *percentage of people with a non-western migration* summing up all remaining countries.

The *average house value (*1000€)* is based on the law for the valuation of immovable properties in the Netherlands. The value is only indicated if at least 50 single values for calculating the average were available.

The *average income per inhabitant (*1000€)* is the arithmetic average personal income per person based on the total population. After an additional inquiry, CBS states that: "An individual's gross income is made up of: income from work, income from enterprise, benefits from income insurance and social benefits (except child allowances). In the variable INK_INW, the mean is calculated for the whole population including the persons without income." The variable describes annual income and is only given if at least 100 single values were available to CBS.

The *percentage of households with the lowest income* is the share of private households that belong to the national 40% of households with the lowest household income.

Lastly, for the variable *percentage of households below or around the social minimum* student households and households with incomplete annual income are not included. The social minimum is the legal minimum which has been determined in political decision-making. This minimum depends on the household e.g. differs between single households and households with children.

For a number of variables, it is important to highlight the standard deviation: Most variables are heterogenous in terms of their values. For example, population density varies a lot with a mean of 2.662 people per square kilometre while there are neighbourhoods which are not at all populated and a neighbourhood with a population density of 28.599 people per square kilometre. These differences make interpretations from the next analysis more difficult and should be kept in mind.

The variable *Neighbourhood size in km2* is not included in the dataset and was instead calculated with the GIS software ArcMap.

Variable	**N**	**Mean**	**SD**	**Median**	**Min**	**Max**
Demographic and auxiliary variables						
Population	12237	1379.23	1977.39	655	0	27650
Percentage of people 65 and older	12237	18.43	10.47	18	0	100
Population density per km2	11736	2662.38	3325.63	1291.5	0	28599
Percentage of people with a western migration background	12237	7.95	6.66	7	0	100
Percentage of people with a non-western migration background	12237	6.19	10	3	0	100
Neighbourhood size in km2	12238	2.86	5.69	0.78	0.01	130.14
Socioeconomic variables						
*Average house value (*1000€)*	9399	240.69	107	220	38	1523
*Average Income per inhabitant (*1000€)*	10110	24.14	5.63	23.2	7.7	84.4
Percentage of households with the lowest income	8357	35.27	14.75	33	2	99
Percentage of households below or around the social minimum	8282	7.12	4.7	6	0	55
Supermarket Access						
Large Supermarket average distance in km	11702	1.61	1.36	1.1	0.1	11.7
Large supermarket average number within 1 km	11702	1.04	1.47	0.5	0	16.1
Large supermarket average number within 3 km	11702	5.99	7.73	3.7	0	88.4
Large supermarket average number within 5 km	11702	12.66	15.41	7.6	0	145.4

Table 1 - Descriptive Statistics of the CBS dataset

In terms of access to supermarkets, the variable *large supermarket average distance in km* gives a first impression of the food accessibility in the Netherlands. In comparison with the study of White et al. (p. 77, 2004), set in the UK, the results in the Netherlands, with a median of 1.1 km and a maximum of 11.7 km average distance to the next large supermarket, are lower (see Appendix 2 – Average Distance to the Next Closest Supermarket). White et al. (2004) found a median distance of 1.8 km and a maximum travel distance of 23.7 km. When only considering these values, the general Dutch food access seems to be superior. However, the data here is on a neighbourhood level while the data by White et al. (2004) is on a much more accurate houshold level.

Table 2 shows the supermarkets chains used in the analysis, including their frequency, after the previously described adjustments.

Chain	**Freq.**	**Per cent**
Albert Heijn	814	24.57
Jumbo	500	15.09
Aldi	464	14.01
Lidl	266	8.03
Coop	209	6.31
Plus	206	6.22
Spar	180	5.43
Emte	110	3.32
Dirk	93	2.81
Deka	81	2.44
Poiesz	70	2.11
Deen	68	2.05
Hoogvliet	62	1.87
Vomar	60	1.81
Jan Linders	59	1.78
Boni	40	1.21
Nettorama	31	0.94

Table 2 - Supermarket Chains used in the analysis

Most stores are operated by Albert Heijn, Jumbo, Aldi, Lidl, Coop and Plus, making up roughly 74% of all chain-supermarkets in the sample. Compared to the original sample containing all grocery stores, including small and “to-go” stores, the chain-supermarkets make up 58% of all grocery stores.

4.2 Measurements of Supermarket Accessibility

Previous results indicate that different measurements for the relatively general term *supermarket accessibility* are necessary to obtain robust results. For example, Helbich et al. (2017) found only moderate associations between accessibility measures. According to the authors, only multiple indicators can frame a comprehensive picture of supermarket accessibility.

The CBS dataset provides two basic measures, which are also the most widely used definitions for food outlet accessibility in the literature (Lamb et al., 2015): Firstly, the proximity measure describes the average distance of all residents in a neighbourhood to the nearest large supermarket (Toelichting Wijk- en Buurtkaart, 2017). The distance is calculated by using the street network. Secondly, the density measure calculates the average number of food outlets for all residents within a given road distance of 1, 3 and 5 kilometres.

Similarly, these measures are again calculated for every neighbourhood using the supermarket data introduced in 3.1: Proximity is thereby defined as the linear distance from each geometric weighted neighbourhood centroid to the next chain-supermarket. Density is calculated as the number of chain-supermarkets within a buffer around the geometric weighted neighbourhood centroid of 1, 3 and 5 kilometres. In the literature, although there is no agreement on distances, a buffer of one kilometre is the most commonly adopted buffer size (Helbich et al., 2017). While Statistic Netherlands uses the average number of supermarkets for all residents in a neighbourhood, the calculated number is not an average and can only be zero or a positive whole number.

Table 3 shows the Pearson correlations between the measurements calculated from the supermarket data and the existing measurements by CBS. In general, the correlation is high as could be expected.

Variables	Correlation	Observations
Proximity	0.7381	10102
Number of supermarkets in 1 km distance	0.8041	10102
Number of supermarkets in 3 km distance	0.9259	10102
Number of supermarkets in 5 km distance	0.9300	10102

Table 3 - Correlation for Supermarket measurements between CBS and Supermarket Data

The small variations may be explained by the different methodologies used: CBS uses the street network and distance from every inhabitant on average while the measurements from the supermarket data were calculated by linear distances and from the centroid of each neighbourhood. Furthermore, CBS defines a large supermarket as a store with several kinds of daily items and a minimum floor space of 150 square meters while the measurements which were calculated make only use of chain-supermarkets.

As already mentioned, various measurements were adopted due to the small correlations between measurements for supermarket accessibility in previous studies. Therefore, these measurements may capture different components of food access. While the correlations between the same measurements of CBS and of the calculated supermarket data are relatively high, Table 4 shows all correlations between the different measurements. The table emphasizes the need for different measurements of food access since many are only weakly to moderately correlated.

	CBS Proximity 1 km	*CBS Density 1 km*	*CBS Density 3 km*	*CBS Density 5 km*	*Supermarket Data Proximity*	*Supermarket Data Density 1 km*	*Supermarket Data Density 3 km*	*Supermarket Data Density 5 km*
CBS Proximity	1							
CBS Density 1 km	-0.5506	1						
CBS Density 3 km	-0.4655	0.7303	1					
CBS Density 5 km	-0.4014	0.6102	0.9161	1				
Supermarket Data Proximity	0.7386	-0.4671	-0.4482	-0.407	1			
Supermarket Data Density 1 km	-0.5217	0.8083	0.7028	0.5803	-0.5468	1		
Supermarket Data Density 3 km	-0.4666	0.6416	0.9219	0.9180	-0.5045	0.6787	1	
Supermarket Data Density 5 km	-0.4011	0.5300	0.8057	0.9269	-0.4452	0.5359	0.8927	1

Table 4 - Correlations between all measurements for food access

Lastly, a measurement of variety was applied which is far less represented in the food desert literature. However, one earlier study set in the Netherlands made use of this measurement (Stelder, 2012). The measurement is supposed to represent consumers' variety of choice in terms of products and prices since various supermarket chains differ in both (Drewnoski et al., 2014). Variety is defined as the number of chain-supermarkets from different Chains within a linear distance of 3 kilometres (similar to Helbich et al. (2017)). This measure can only take four different values: 1 if all three nearest supermarkets are operated by the same chain, 2 if only two of the nearest supermarkets are operated by the same chain & 3 if all three nearest supermarkets are operated by different chains. In addition, the measure can take a value of 1 if only supermarkets by one Chain are in a distance of 3 kilometres from the neighbourhood centroid. For example, in many rural areas there is only one supermarket in such a distance available. Similarly, the measurement can take a value of 2 if only supermarkets by two different chains are in a distance of 3 kilometres from the neighbourhood's centroid. If no supermarket is in a distance of 3 kilometres, the measurement takes the value 0. For the statistical analysis, different thresholds then 3 kilometres were also used.

	Freq.	**Per cent**
No supermarkets in a 3 km Radius from the Neighbourhood Centroid	2030	16.45
All three nearest supermarkets in a 3 km Radius from the Neighbourhood Centroid are operated by the same Chain (Includes cases where there are only one or two supermarkets)	1697	13.75
The three nearest supermarkets in a 3 km Radius from the Neighbourhood Centroid are operated by two different Chains (includes cases where there are only two supermarkets)	2827	22.91
All three nearest supermarkets from the Neighbourhood Centroid are operated by different Chains	5786	46.89
Total	11,118	100

Table 5 - Variety measurement

Table 5 shows the distribution of the variety measurement. The most problematic cases, in which there is either no supermarket in a 3 kilometres radius or only supermarkets by one chain, affect 30.2% of all neighbourhoods. This corresponds to 2.363.065 people of the 16.778.365 people in the dataset. Therefore, since the share of the population of these neighbourhoods is smaller than the share of the number of neighbourhoods, mostly neighbourhoods with a smaller population are affected. However, these numbers should be taken with caution since about 20% of population data is missing.

Possible measurements of supermarket accessibility which were not carried out here include travel time to the nearest food outlet or a binary outcome measurement of presence or absence of a supermarket in a certain buffer (Charreire et al., 2010).

4.3 Example for measuring Supermarket Accessibility

The previously introduced measurements, proximity, density and variability, might seem abstract. Therefore, this chapter applies these measurements for a neighbourhood in Groningen: *Helpman*. The neighbourhood is located in the south of Groningen and has a population density of 10.425 people per square kilometre. While the neighbourhood is not in the centre of Groningen, the population density is still relatively high. As a comparison, the highest population density in Groningen is 16.016 people per square kilometre in the *Binnenstad-West*.

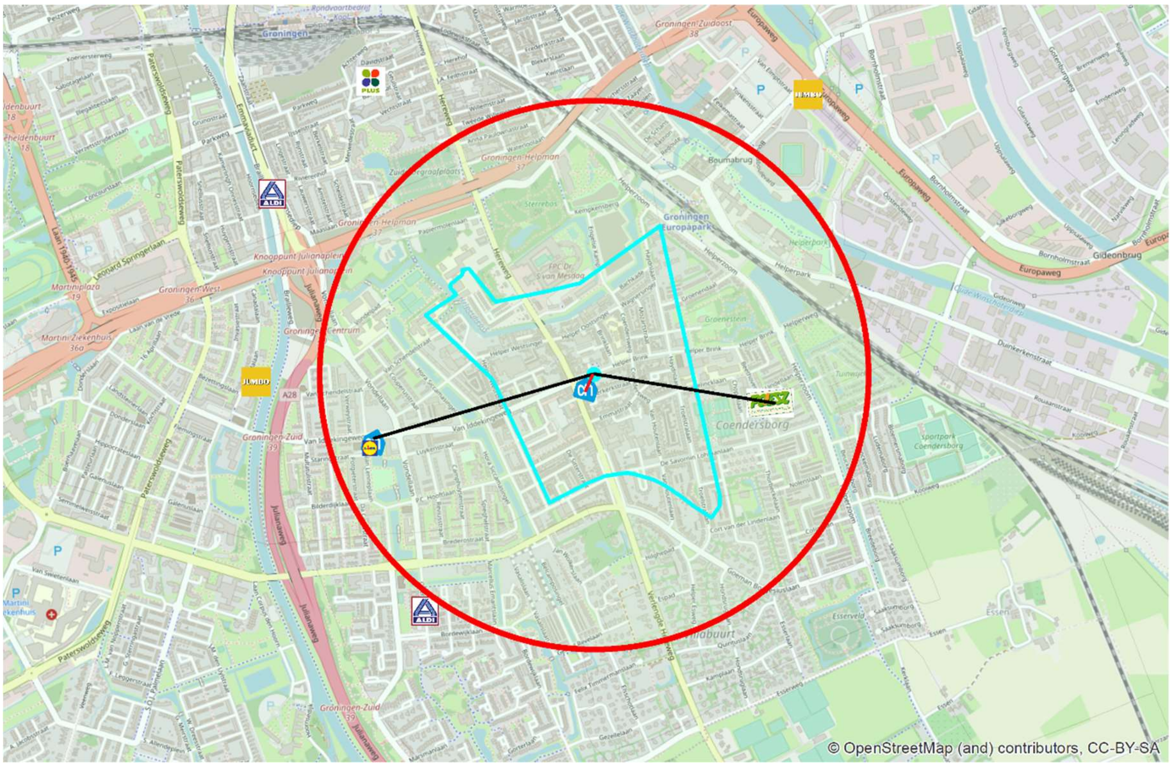

Figure 1 - Map of Helpman (Groningen) with the three introduced measurements

The Figure 1 shows a map of Helpman marked by a light blue border () and its centroid as a light blue point (). In addition, the three different measurements are plotted.

First, the red line () represents the proximity indicator. In this case the proximity from the neighbourhood centroid to the closest supermarket is very low with only 60 meters.

Second, the red circle () shows a buffer of 1 kilometre from the neighbourhood centroid. The density measurement, therefore, takes a value of 4 since the buffer includes four supermarkets: Two Albert Heijn, one Lidl and one Poisez. The buffer of one kilometre already visualises how large such a distance, especially in urban areas, is. Approximately there are about 30.000 people living even in this smallest buffer size. The main attention should therefore be on the distance for the buffer of 1 kilometre when conducting the further analysis. This buffer has already been identified as the most common in previous research (Helbich et al., 2017).

Lastly, the three nearest supermarkets are identified with the red line (), which represents the closest supermarket, and the two black lines () which connect the second and third closest supermarket to the neighbourhood centroid. These stores are operated by either Albert Heijn or by Poisez. Thus, since two different chains operate the three nearest stores, the variability measurement takes a value of 2. However, at this point, a limitation should be highlighted: The Albert Heijn which is operating in the east is only a few meters (~30m) apart from a Lidl. Therefore, the variability of 2 does not match the reality: The population of Helpman can nearly as easily reach three different chains.

4.4 Possible Data Issues

Generally, the sample should be representative of the Dutch population since the data by CBS as well as the supermarket data is for the whole of the Netherlands. However, there are a few data issues which will be discussed in this chapter. These issues might make a generalisation of the results difficult without considering some limitations.

4.4.1 Incomplete Supermarket data

The supermarket data does not contain all supermarkets in the Netherlands. For example, in the sample, there are 206 locations of the supermarket chain Plus. However, according to Plus' official site, the chain operates 260 stores in the Netherlands (Plus, 2019). While that difference is not as big for all other chains, the number of stores included in the analysis is always smaller than the number of stores advertised on the official sites.

When choosing an approach where the spatial distribution of supermarkets is mapped with socio-demographic data, one would identify far too many areas with bad access to food as a large number of supermarkets is missing in the sample.

However, this study is focusing on analysing the statistical relationship between socioeconomic factors and *distance to* or the *number of* supermarkets. In all conducted models the distances to supermarkets may be overestimated and the numbers of supermarkets may be underestimated. However, there is no reason to assume that these missing values are in any way correlated with any other variable and, thus, should only reduce the efficiency and not bias the estimates of the coefficients (Jakobsen & Mehmetoglu, 2016).

4.4.2 Missing Data in the CBS dataset

All variables in the CBS dataset have missing data points for multiple neighbourhoods. Table 6 highlights these missing values. While for the demographic variables and the variables concerning supermarket proximity the percentage of missing values is at about 5% or less, the number of missing values is much higher for all socioeconomic variables. Especially values concerning the identification of low-income areas are missing in about one-third of all neighbourhoods. Since the main goal of this analysis is to identify the relationship between socioeconomic variables and supermarket proximity, this may become a major issue.

Variable	N	Missing	Percentage of missing values	Mean
Demographic and auxiliary variables				
Population	12237	1	0.01	1379.23
Area	12238	0	0	2.86
Percentage of people 65 and older	12237	1	0.01	18.43
Population density per km2	11736	502	4.1	2662.38
Percentage of people with a western migration background	12237	1	0.01	7.95
Percentage of people with a non-western migration background	12237	1	0.01	6.19
Socioeconomic variables				
*Average house value (*1.000€)*	9399	2839	23.2	240.69
*Average Income per inhabitant (*1.000€)*	10110	2128	17.39	24.14
Percentage of households with the lowest income	8357	3881	31.71	35.27
Percentage of households below or around social minimum	8282	3956	32.33	7.12
Supermarket Access				
Large Supermarket average distance in km	11702	536	4.38	1.61
large supermarket average number within 1 km	11702	536	4.38	1.04
large supermarket average number within 3 km	11702	536	4.38	5.99
large supermarket average number within 5 km	11702	536	4.38	12.66

Table 6 - Missing data in the CBS dataset

In any case, the efficiency of the coefficients will be reduced since the number of usable cases is reduced (Schafer & Graham, 2002). According to Schafer & Graham (2002), whether missing values bias the coefficients depends on the randomness of the missing values. For most variables, CBS only states that these datapoints were either unknown, insufficiently reliable or kept secret (Toelichting Wijk- en Buurtkaart, 2017). To conduct the analysis, the missing values are assumed to be unrelated to their value and all other variables. Listwise Deletion will be used when necessary to deal with the missing data: Each case that has a missing value will be dropped from the analysis.

The biggest problem is connected to both the income and the housing value variable: CBS states that these variables are only available for neighbourhoods with at least 100 inhabitants or 50 houses (Toelichting Wijk- en Buurtkaart, 2017). Thus, all smaller neighbourhoods are not used in the analysis which may alter the results significantly. This may bias the results since these values are not missing at random.

4.4.3 Multicollinearity

A number of variables are very similar in terms of what these variables are measuring. For example, there are variables for the percentage of households with the lowest income, the percentage of households below or around the social minimum and for the general income. Since these variables are very similar, multicollinearity may be an issue when conducting any analyses. Multicollinearity exists when two or more variables in a regression model are highly correlated (Jakobsen & Mehmetoglu, 2016).

	Population Density per km2	**People aged 65 and older**	**Percentage of people with a western-migration background**	**Percentage of people with a non-western migration background**	**Average Income per Inhabitant**	**Average house value(*1.000€)**	**Percentage of household with the lowest income**	**Percentage of households below or around the social minimum**
Population Density per km2	1							
People aged 65 and older	-0.1986	1						
Percentage of people with a western-migration background	0.3616	0.0298	1					
Percentage of people with a non-western migration background	0.5828	-0.2247	0.333	1				
*Average Income per Inhabitant (*1.000€)*	-0.1177	0.1834	0.1831	-0.2718	1			
*Average house value(*1.000€)*	-0.3754	0.0635	-0.0914	-0.3534	**0.7257**	1		
Percentage of household with the lowest income	0.3937	0.1399	0.3576	0.494	-0.5094	-0.5579	1	
Percentage of households below or around the social minimum	0.3716	-0.0544	0.3488	0.6525	-0.4386	-0.3864	**0.8162**	1

Table 7 - Correlation between all explanatory variables

Table 7 shows that most correlations are relatively low and therefore manageable. However, the correlation between the *percentage of households below or around the social minimum* and the *percentage of households with the lowest income* as well as the correlation between the *average house value* and the *average income per inhabitant* require further investigation. Both correlations display a value higher than 0.7 and, thus, could become problematic for the analysis.

Variable	**VIF**	**1/VIF**
Percentage of households with the lowest income	4.95	0.20201
Percentage of households below or around the social minimum	4.66	0.214745
*Average income per inhabitant (*1.000€)*	3.51	0.285302
*Average house value (*1.000€)*	3.02	0.331254
Percentage of people with a non-western migration background	2.44	0.409099
Population Density per km2	1.91	0.523679
Percentage of people with a western-migration background	1.63	0.614915
People aged 65 and older	1.39	0.71732

Table 8 - Variance Inflation Factors for explanatory variables

Recent literature recommends the use of Variance Inflation Factors (VIF) which are shown in Table 8 (Daoud, 2017). The VIF is a tool to measure and quantify how much the variance in a regression is inflated. Usually, a VIF higher than 5 is seen as very problematic. In this case the *percentage of households with the lowest income* is close to this threshold. Thus, the variable is dropped from the analysis. After dropping the variable, the VIF of the *percentage of households below or around the social minimum* decreases to an unproblematic 2.38.

5. Graphical visualisation of Supermarket Accessibility

Before a statistical analysis is conducted, the hypotheses are tested graphically. For these investigations, cumulative distribution functions are used. Cumulative distribution functions give the fraction of the outputs that are less than or equal to a given value of, in this case, distance to the next supermarket (Xue et al., 2009).

To compare the access to supermarkets using cumulative distribution functions, the Dutch population was separated into three urbanicities: Urban areas, urban clusters and rural areas.

Urban areas are defined as areas with a population density of at least 1.500 inhabitants per square kilometre, urban clusters are neighbourhoods with a density between 500 and 1.500 inhabitants per square kilometre and rural areas are neighbourhoods with a density less than 500 inhabitants per square kilometre. However, these definitions are mostly arbitrary while being loosely based on the data and previous literature (Ver Ploeg et al., 2009).

Figure 2 shows the distance to supermarkets by urbanicity, using only the average distance to the next large supermarket from the CBS dataset. The same graphs which are presented in this chapter were also created using the centroid distance to the closest chain-supermarket. These can be found in Appendix 8.3.

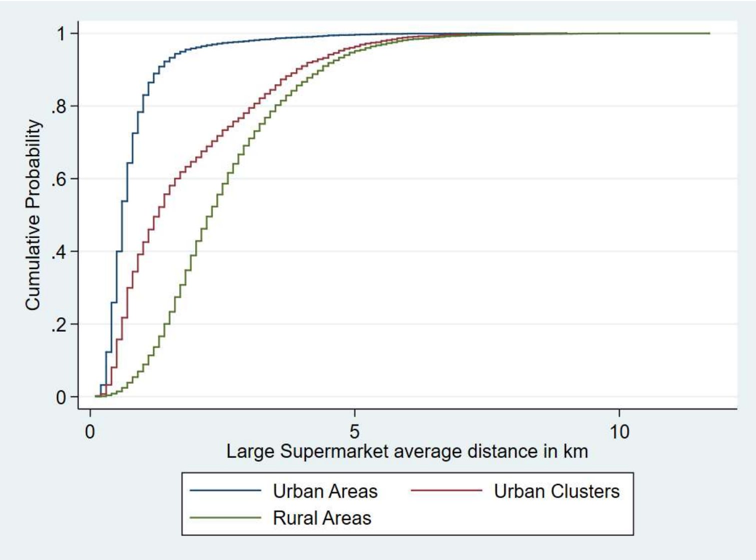

Figure 2 - Distance to Supermarket by Urbanicity

As could be expected, the average distance to a large supermarket in Urban Areas is lower than in Rural areas. Furthermore, 99% of all Rural areas have an average distance to the next large supermarkets of less than 6.6 kilometres in the dataset. For the US, Ver Ploeg et al. (p. 135, 2009) found that not even 50% of the rural population live in a distance of 6 kilometres to a large supermarket. However, the study differs significantly in many parts:

Areas are defined under other conditions as rural and the US is not comparable to the Netherlands in terms of general population density.

5.1 Supermarket Access for Low-Income groups

To assess the impact of income differences on supermarket accessibility the dataset was divided into three groups: low-income, medium-income & high-income.

Low-income neighbourhoods were defined as neighbourhoods within the bottom 20% of the average income per inhabitant distribution while high-income neighbourhoods were defined as neighbourhoods in the top 20% of the average income per inhabitant distribution. The remaining neighbourhoods are medium-income neighbourhoods.

Figure 3 shows the income distribution marking the different income groups.

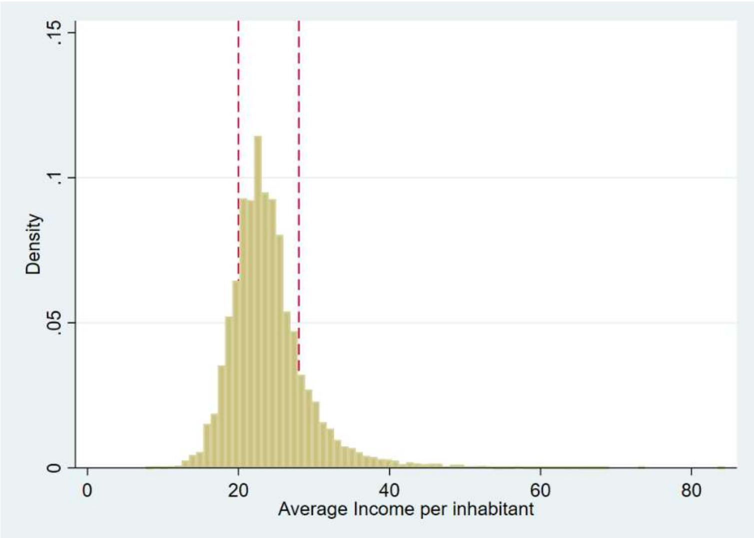

Figure 3 - Income Distribution

To adjust the analysis for differences in urbanicity, Figure 4, Figure 5 and Figure 6 show the average distances to a large supermarket for the various income groups for each Urbanicity type in cumulative distribution functions.

The differences in *Urban Areas* in terms of income are rather small. There are a number of low-income neighbourhoods which are closer to a large supermarket than high-income areas. However, while there are nearly no high-income neighbourhoods further than 3 kilometres from a supermarket, a small fraction of low- and medium-income neighbourhoods are.

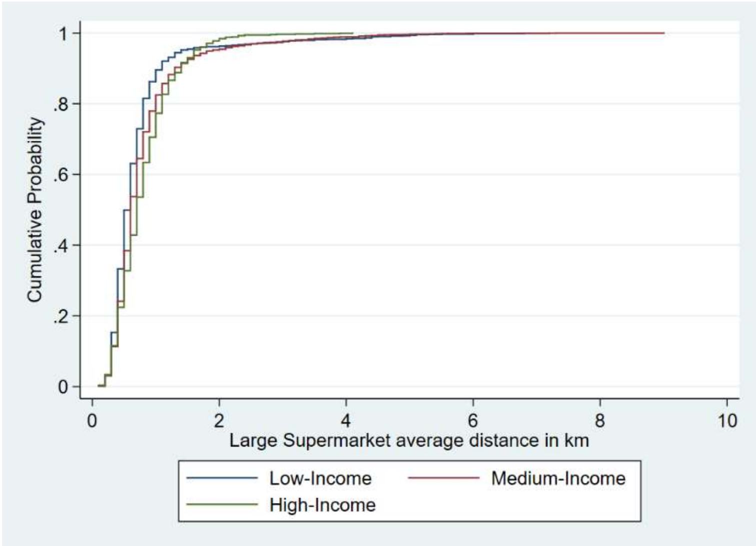

Figure 4 - Distance to Supermarket by Income Groups in Urban Areas

In *Urban Clusters,* these differences are more pronounced. There are clearly more low- and medium-income located further from a supermarket than there are high-income neighbourhoods. However, these differences only occur for an average distance between approximately 2 and 5 kilometres. Neighbourhoods which are closer to a large supermarket are mixed in their income level with a slight advantage of low-income groups being the closest for this group.

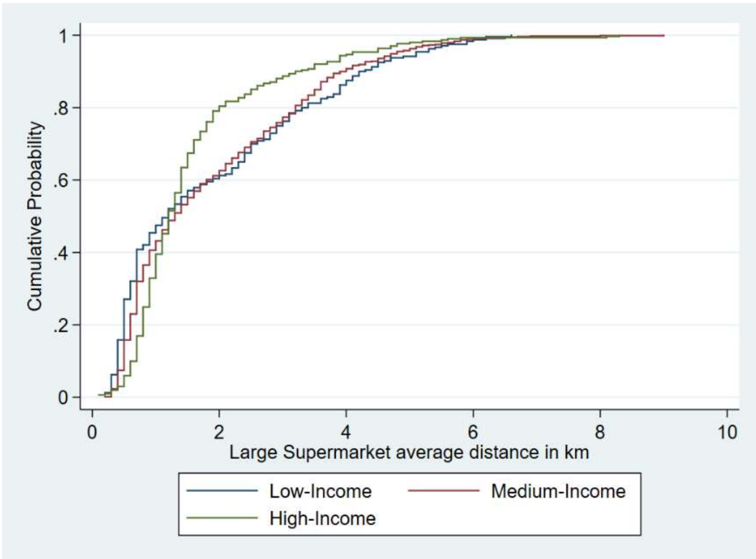

Figure 5 - Distance to Supermarket by Income Groups in Urban Clusters

In rural areas the differences are not particularly large but clear for all distances: High-income neighbourhoods are closer to supermarkets than medium-income neighbourhoods. Moreover, medium-income neighbourhoods are closer to supermarkets than low-income neighbourhoods.

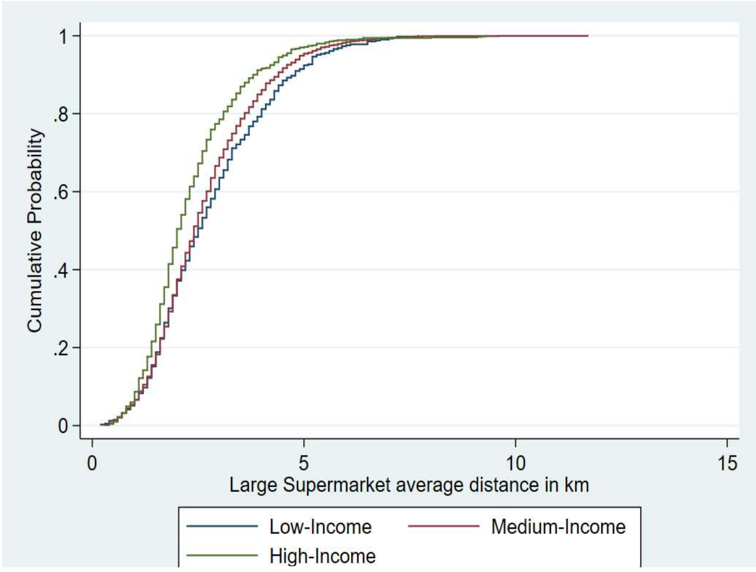

Figure 6 - Distance to Supermarket by Income Groups in Rural Areas

All of these results are backed up by the proximity indicator which was calculated with the locations of the supermarkets.

In conclusion, it seems like low-income neighbourhoods are, mostly, closer to the next large supermarket in densely populated areas while in less densely populated Urban Clusters and Rural areas low-income neighbourhoods are further from the next large supermarket. Since the distances to supermarkets are significantly longer in rural compared to urban areas, rural areas may be the most problematic areas for low-income households.

5.2 Supermarket Access for Immigrants

For the graphical analysis of the relationship between supermarket access and areas with high shares of minorities one additional indicators is constructed:

The total percentage of immigrants is estimated by summing up the western- and non-western-migration. Afterwards, neighbourhoods in the top 20% of total immigrant percentage which corresponds to areas with at least 22% of immigrants, as shown by Figure 7, are defined as areas with a high share of immigrants.

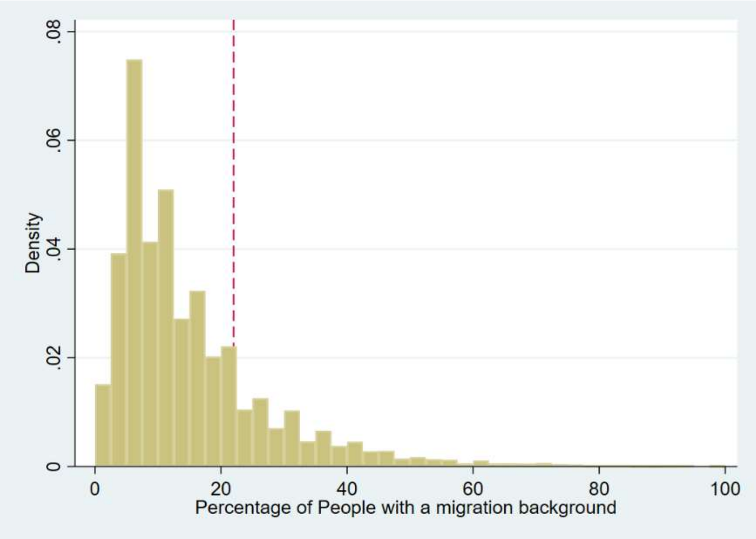

Figure 7 - Total Immigrant Percentage Distribution

Like before, the relationship is analysed for Urban Areas, Urban Clusters and Rural areas.

Firstly, Figure 8 and Figure 9 show the relationship between average distance to supermarkets and areas with high percentages of immigrants in Urban Areas and Urban Clusters.

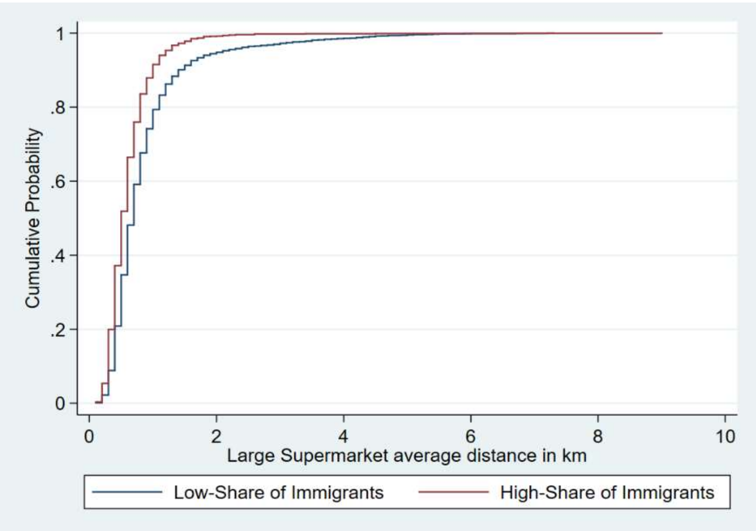

Figure 8 - Distance to Supermarket by Migration Percentage in Urban Areas

When controlling for urbanicity, high-share immigrant areas are closer to supermarkets compared to low-share immigrant areas in Urban Areas. The same results are obtained in Urban Clusters. Thus, there seems to be no discrimination happening in Urban Areas and Urban Clusters in terms of supermarket access.

Lastly, Figure 10 displays the relationship in rural neighbourhoods. While the results are similar for neighbourhoods located closer than 4.45 kilometres to a supermarket, neighbourhoods located even further from a supermarket are more often high-share immigrants neighbourhoods. Like in the previous chapter, the most problematic neighbourhoods in terms of supermarket access for minority groups, seem to be rural areas far away from a supermarket.

The results are, approximately, the same when using the calculated distance measure for supermarket access (see Appendix 8.3).

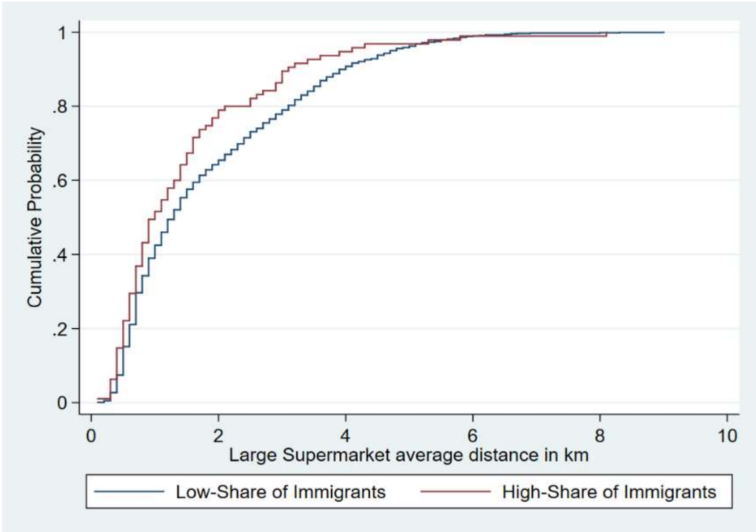

Figure 9 - Distance to Supermarket by Migration Percentage in Urban Clusters

All in all, in opposition to the hypothesis set up in chapter 3.2, areas with percentages of migrants seem mostly closer to the next supermarket compared to areas with a lower percentage of migrants. One exception are rural areas: Some of the neighbourhoods with a high share of immigrants are especially far from the next supermarket.

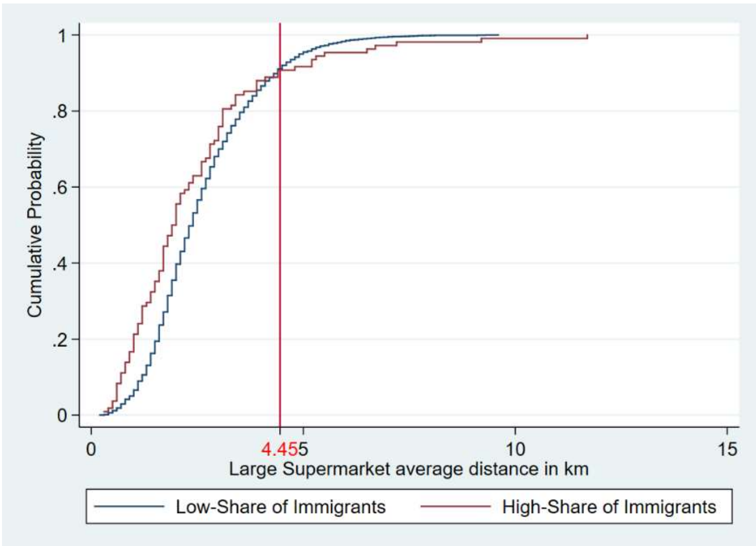

Figure 10 - Distance to Supermarket by Migration Percentage in Rural Areas

5.3 Supermarket Access for Elderly People

Lastly, access for elderly people is analysed graphically. Similarly, as for migrants, an area is defined as having a high share of elderly people if it is in the top 20% in terms of the percentage of people aged 65 or older. This equals areas with at least 24% of people aged 65 and older which is visualized by Figure 11.

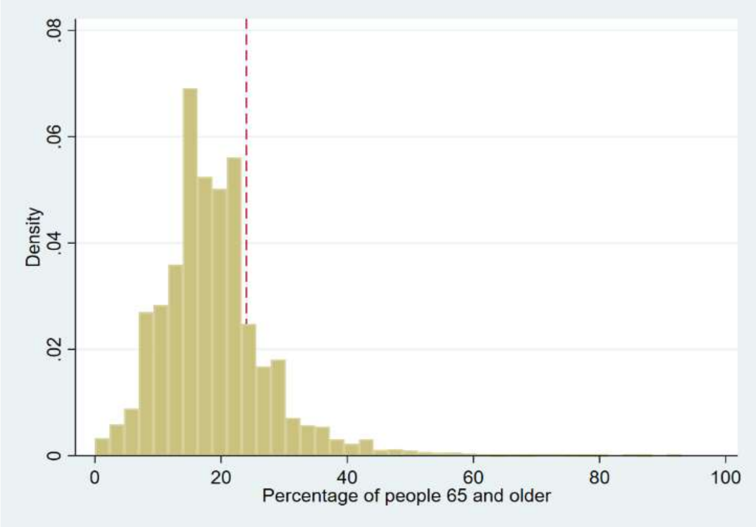

Figure 11 – Distribution of the Percentage of people aged 65 and older

Again, the relationship is analysed for each urbanicity type: Urban Areas, Urban Cluster and Rural Areas.

Figure 12 displays the relationship between areas with a high amount of people aged 65 and older and average distance to supermarkets in Urban Areas. Elderly people seem to locate closer to supermarkets than younger people in Urban Areas.

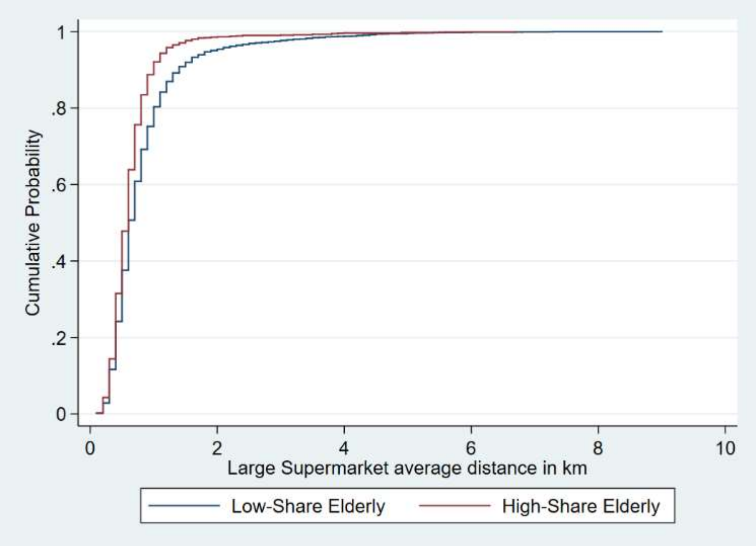

Figure 12- Distance to Supermarket by Elderly Percentage in Urban Areas

Similarly, Figure 13 shows the same relationship in Urban Clusters. In these areas, which are not as densely populated as Urban areas, the relationship is even more pronounced: About 80% of neighbourhoods in Urban Clusters with a high share of elderly people are within 2 kilometres of a supermarket. In comparison, only about 60% of neighbourhoods in Urban Clusters with a low share of elderly people are within 2 kilometres of a supermarket.

Lastly, Figure 14 highlights this relationship in rural areas. The results are similar: In general, rural neighbourhoods with high shares of elderly people are closer to supermarkets compared to neighbourhoods with low shares of elderly people.

All in all, elderly people in the Netherlands seem to locate closer to supermarkets than younger people. Therefore, the graphical analysis confirms the previously established hypothesis.

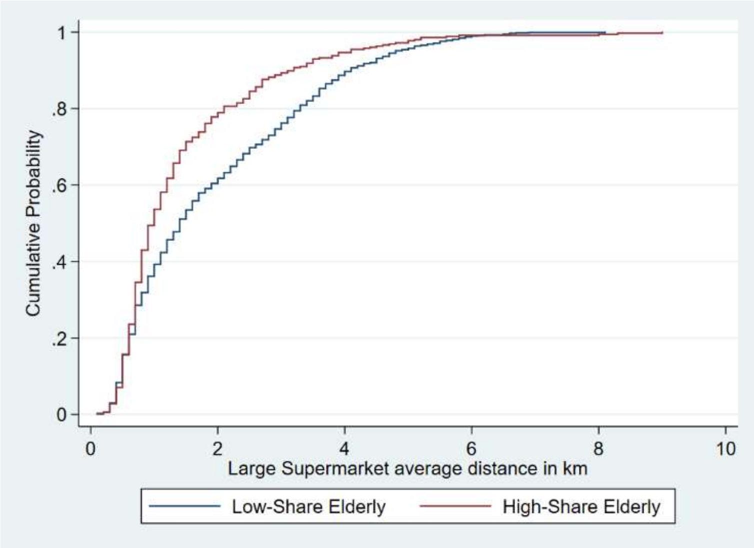

Figure 13 - Distance to Supermarket by Elderly Percentage in Urban Clusters

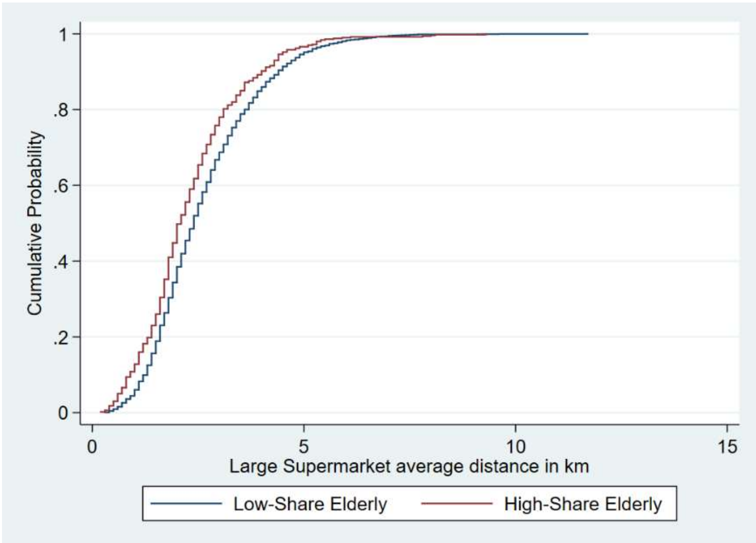

Figure 14 - Distance to Supermarket by Elderly Percentage in Rural Areas

6. Main Statistical Analysis

The graphical illustrations in the previous chapter give a first impression about the relationship between socioeconomic indicators and supermarket access in the Netherlands. However, since these distributions only examine one indicator at a time, possible underlying mechanisms are missed. For example, this may be the case for elderly people: The graphical analysis finds that the elderly, generally, live in neighbourhoods which are closer to supermarkets. Nevertheless, the true underlying factor may be that elderly people have a higher income than younger people. Since this underlying characteristic is ignored, wrong conclusions could be drawn. That is why this chapter investigates the impact of socioeconomic indicators on a number of supermarket access measurements with various regressions.

The different access measurements are used to test the robustness of the results. If the results do not change significantly, broader conclusions may be drawn from the analysis. For a better comparison between the models, the various measures by CBS were transformed from kilometres to metres.

To determine which variables are taken into each model, the forward stepwise selection procedure is used with a threshold-significance level of 10%. The selection procedure starts with no candidate variables in the model and adds one by one the most significant variable (Stata, 2019). This goes on until no variable can be added with a significance level lower than 10%. All explanatory variables introduced in 4.1 are used except for the *Percentage of household with the lowest income* due to the issues of multicollinearity described in 4.4.3.

6.1 Spatial Autocorrelation

When analysing the availability and accessibility of supermarkets in neighbourhoods, it should be acknowledged that the data is spatial. Ignoring this characteristic when determining the right statistical approaches to employ in the analysis may lead to biased results (Lamb et al., 2015). In this context, positive spatial correlation means that areas near one another have more attributes in common with each other than with areas further away (Cliff & Ord, 1995).

Many studies assume that the residuals are independently distributed. However, this assumption should always be verified when dealing with spatial data (Lamb et al., 2015). If the spatial autocorrelation is not taken into account, the coefficients may be biased (Lamb et al., 2012).

The most common test for spatial autocorrelation used in the analysis of accessibility measures of food outlets is Moran's I (Lamb et al., 2015; Moran, 1950). The test examines the null hypothesis of no spatial autocorrelation between data zones. To define the spatial relationship an inverse distance matrix was used. The inverse distances were calculated with the formula:

$$w_{ij} = \frac{1}{d_{ij}^{1}}$$

w_{ij} gives a weight to the spatial relationship between neighbourhood centroid i and neighbourhood centroid j. The distances d_{ij} were calculated from neighbourhood centroid i to neighbourhood centroid j. A threshold was introduced by ArcMap automatically. Thus, only relatively close neighbourhoods were assumed to affect each other.

Necessary adjustments such as dropping missing cases were performed with Stata. All non-spatial regression models were calculated in Stata while spatial models were calculated with a script in MATLAB (LeSage, 2010).

The residuals will be tested for spatial autocorrelation after running each regression for Proximity, Density and Variety measures. If this is the case a spatial model will be adopted instead of a non-spatial model.

6.2 Proximity to Supermarkets

The models estimating socioeconomic indicators on Proximity in Meters to the closest supermarket are calculated by OLS as recommended by Lamb et al. (2015).

Table 9 shows the results for both Proximity variables: The CBS measure (model 1), which measures the average distance, and the calculated measure to the next chain-supermarket from each neighbourhood centroid (model 2).

	CBS Data (model 1)	Supermarket Data (model 2)	Spatial CBS Data model (model 3)	Spatial Supermarket Data (model 4)
*Average Income per Inhabitant (*1.000€)*	-35.0726*** (3.591494)	-47.57514*** (4.231115)	-43.0980112*** (3.673805)	-41.5651346*** (3.445535)
Percentage of households below or around social minimum	17.25654 *** (3.556841)	19.24186*** (4.181026)	13.0705031** (3.419940)	4.3535415 (3.059496)
*Average house value (*1.000€)*	2.924446*** (0.1822201)	2.68584*** (0.2147891)	3.8026186*** (0.185765)	2.7870342*** (0.171067)
Percentage of people 65 and older	-21.69721*** (1.299545)	-20.23388*** (1.527708)	-17.6789640*** (1.198589)	-11.5563172*** (1.065084)
Percentage of people with a non-western migration background	-14.80991 *** (1.612688)	-20.44993*** (1.898118)	-18.6276660*** (1.775194)	-16.3853071*** (1.692473)
Percentage of people with a western migration background	-10.13295*** (2.608573)	-16.0205*** (3.074465)	-6.9492262* (3.005863)	-6.4593638* (2.939249)
Population density per km2	-0.1261352*** (0.0041921)	-0.1407778*** (0.0049415)	-0.0930433*** (0.004500)	-0.0743889*** (0.004199)
Constant	2370.578*** (66.97272)	2790.366*** (78.83183)	2281.8304675*** (78.38146)	2588.019*** (84.10604)
Rho of interaction effects among the error terms (λ)			0.5680*** (0.01145180428)	0.7900*** (0.006688)
R-squared	0.3018	0.2821	0.4861	0.6921
Adjusted R-squared	0.3012	0.2815	0.4857	0.6918
Sample Size	8253	8258	8253	8258
Moran's I of Residuals	0.097632	0.102190	0.077640	0.113588
p- value for Moran's I of Residuals	0.000000	0.000000	0.000000	0.000000

Note: *Numbers in parentheses are standard errors*
****p<0.001, **p<0.01, *p<0.05*
Dependent Variable: Distance to next Supermarket in Meters

Table 9 - Regressions with Proximity to closest Supermarket as the dependent variable

Since both models suffer under significant spatial autocorrelation among the error terms, one spatial econometric model for each measurement is introduced. Spatial econometric models deal with interaction effects among neighbourhoods in a number of different ways (Elhorst, 2014). In this case, *spatial interaction effects among the error term* are added to the model:

$$Y = X\beta + u$$

$$with: u = \lambda W u + \epsilon$$

In the model Y is the dependent variable, X the matrix of independent variables, β the vector of coefficients, λ the coefficient for the spatial interaction effects among the error term and W the spatial weight matrix. The spatial weight matrix W differs from the one used for assessing Moran's I due to incompatibilities between ArcMap and MATLAB. For the spatial econometric model, W is again defined by the inverse distances described earlier, but instead of using a distance threshold, the maximum number of neighbours is limited to five.

The spatial model shown here, which is also called Spatial Error Model, is used since it is likely that unobserved factors in some areas may influence the results. For example, the attributes of inner-city neighbourhoods used in the analysis may not solely determine the distance to the next supermarkets. Since many people commute to these neighbourhoods for work, the demand may be higher than the characteristics of the neighbourhood suggest. By incorporating spatial interaction effects among the error term, this impact should be controlled for.

All in all, no signs change in the spatial models. Nonetheless, the error terms are still spatially dependent. This may be the case because of the use of a different weights matrix for calculating Moran's I. Most importantly, however, some variables become insignificant.

In all models, the higher the average neighbourhood income is, the closer is the neighbourhood to a supermarket. However, higher housing price values are related to larger distances to a supermarket. In addition, if a neighbourhood has a higher percentage of households living around the social minimum, the distance to a supermarket is also higher, though this variable is not significant in all models. Adding spatial effects reduces the significance clearly. All these observations are in line with the hypotheses in 3: The demand in high-income areas may be significantly higher than in low-income areas which is why supermarkets prefer to locate in these areas. Nevertheless, higher housing values result in a reduction of profit, which results in an effect in the opposite direction.

Also, this effect does not appear to be small. For example, neighbourhoods with a 1.000€ higher average yearly income as described in 4.1, are located on average about 43 meters closer to the next large supermarket.

Both migration variables have the exact opposite effect compared to the theoretical ideas brought forward in 3. Thus, it does not seem like migrants live further from supermarkets than natives. However, *why* this effect runs the other way remains uncertain.

6.3 Density of Supermarkets

For models estimating the density of supermarkets in a certain radius, OLS is not a suitable estimation method. The high number of 0s, 1s and 2s as well as the lack of negative numbers would distort the results. Thus, Lamb et al. (2015) recommend the use of count models. For the analysis, Poisson estimations are used to take care of the aforementioned problems. Poisson estimates have two particular characteristics (Hoffman, 2016): First, Poisson regression is estimated under the assumption that the mean equals the variance of the dependent variable. For the dependent variables used in the estimations in this chapter, this assumption is mostly satisfied. Second, the events, here the number of supermarkets, have to be independent. Since neighbourhoods close to each other may have the same supermarkets in a specific radius, the density measure may be more similar for neighbourhoods in close proximity. Therefore, this assumption may be violated and should be kept in mind when interpreting the results.

In addition, no script incorporating spatial effects for Poisson regression could be found. That is why no spatial models are used in this chapter. Again, this should be kept in mind when interpreting the results, since some indicators became insignificant after incorporating spatial effects in the previous chapter.

As examined in chapter 4.3, when giving an example for the various measurements, a buffer of one kilometre may be the most meaningful buffer size.

	Supermarket Data 1 km radius (model 1)	CBS Data 1 km radius (model 2)	Supermarket Data 3 km radius (model 3)	CBS Data 3 km radius (model 4)	Supermarket Data 5 km radius (model 5)	CBS Data 5 km radius (model 6)
*Average Income per Inhabitant (*1.000€)*	0.0608504*** (0.0028412)	0.0552877*** (0.0032051)	0.0832714*** (0.0012041)	0.0820919*** (0.0012606)	0.0666769*** (0.0008393)	0.0782639*** (0.0008528)
Percentage of households below or around social minimum	0.0608504*** (0.0024433)	0.0313773*** (0.0022234)	0.0291063*** (0.0010947)	0.0385386*** (0.0011062)	0.0170638*** (0.0008048)	0.0297014*** 0.0007936)
*Average house value (*1.000€)*	-0.004926*** (0.0001828)	-0.0047081*** (0.0002053)	-0.0038462*** (0.0000741)	-0.0037078*** (0.0000775)	-0.0021947*** (0.0000479)	-0.0028056*** (0.0000502)
Percentage of people 65 and older	0.0125046*** (0.0008665)	0.0152765*** (0.0009295)	-0.0026449*** (0.0004645)	-0.0013502** (0.000483)	-0.0038838*** (0.0003362)	-0.0032758*** (0.0003465)
Percentage of people with a non-western migration background	0.0027886** (.0009795)	*Not included in stepwise regression – Insignificant at α=10%*	0.0130891*** (0.0004239)	0.0113533*** (0.0004296)	0.01845*** (0.0003054)	0.0181769*** (0.0003003)
Percentage of people with a western migration background	0.0139996*** (0.001773)	0.0144775*** (0.0144775)	0.0253203*** (0.0007627)	0.0261071*** (0.0007859)	0.0275126*** (0.0005383)	0.0313971*** (0.0005322)
Population density per km2	0.0000987*** (2.18e-06)	0.0001009*** (2.25e-06)	0.0000701*** (1.01e-06)	0.0000787*** (1.01e-06)	0.0000586*** (7.55e-07)	0.0000664*** (7.30e-07)
Constant	-0.8670131*** (0.054547)	-1.163714*** (0.0616101)	-0.0028308 (0.0221917)	-0.2338853*** (0.0232153)	0.8383264*** (0.0151555)	0.4139942*** (0.015529)
Pseudo R-squared	0.2074	0.2131	0.3896	0.4203	0.4134	0.4920
Sample Size	8258	8253	8258	8253	8258	8253

Note: *Numbers in parentheses are standard errors*
****p<0.001, **p<0.01, *p<0.05*
Dependent Variable: Number of Supermarkets in various radiuses

Table 10 - Regressions with Density of Supermarkets as the dependent variable

Table 10 shows the results of the Poisson regression. For most variables, the sign, as well as the significance, stay the same throughout all models. However, there are two notable exceptions: First, the percentage of non-western immigrants is not included in model 2 due to insignificance. Second, the sign of the coefficient for the percentage of people over 65 changes for different radiuses. While the percentage of elderly people in a neighbourhood is positively correlated with the number of supermarkets in a distance of 1 kilometre, it is negatively correlated with the number of supermarkets in 3 or 5 kilometres. One possible explanation may be that some elderly people, who are able to afford this, locate very close to supermarkets. Nonetheless, a number of elderly people, who cannot afford this, are located further from supermarkets than the average person. However, if this is truly the case needs further research.

Since the different dependent variables proximity and density are used to double check the results, changes compared to the results of 6.2 are of special interest. To compare the results one should note the different character of the dependent variable: While in chapter 6.2 positive coefficients indicated worse access, here, positive coefficients indicate better access.

Only one variable has a significantly different impact compared to 6.2. Higher percentages of households around or below the social minimum are correlated with a higher number of supermarkets. In the proximity models, this variable was not always significant. In addition, the results are in contrast with the theory explained in 3.1. Thus, no clear conclusions can be drawn for the impact of the percentages of households around or below the social minimum.

6.4 Variety of Supermarkets

As the Variety measurement can only take values of 0, 1, 2 and 3 Poisson regression is, again, a suitable choice for conducting the analysis. Both assumptions of the Poisson regression are similarly satisfied as described in the previous chapter. Interpretations should be done with care due to the missing inclusion of spatial effects. Table 11 shows the results of the regression. In addition to the threshold of 3 kilometres (model 1) described in 4.2, two other models were added with a threshold of 5 kilometres (model 2) and of 8 kilometres (model 3).

	Supermarket Data, threshold at 3 km (model 1)	Supermarket Data, threshold at 5 km (model 2)	Supermarket Data, threshold at 8 km (model 3)
*Average Income per Inhabitant (*1.000€)*	0.0152389*** (0.0024835)	0.0025203 (0.0014263)	-0.0021531 (0.0012702)
Percentage of households below or around social minimum	-0.0082509** (0.0024564)	-0.0047236* (0.0021753)	*Not included in stepwise regression – Insignificant at α=10%*
*Average house value (*1.000€)*	-0.0008025*** (0.0001329)	*Not included in stepwise regression – Insignificant at α=10%*	*Not included in stepwise regression – Insignificant at α=10%*
Percentage of people 65 and older	0.0048819*** (0.0008636)	0.0024815** (0.0008129)	*Not included in stepwise regression – Insignificant at α=10%*
Percentage of people with a non-western migration background	0.0087065*** (0.001021)	0.0046395*** (0.0010111)	*Not included in stepwise regression – Insignificant at α=10%*
Percentage of people with a western migration background	0.0071803*** (0.0017327)	*Not included in stepwise regression – Insignificant at α=10%*	-0.0023978 (0.0013576)
Population density per km2	0.0000269*** (2.63e-06)	0.0000117*** 2.42e-06	*Not included in stepwise regression – Insignificant at α=10%*
Constant	0.313629*** (0.0459078)	0.747932*** (0.0419504)	1.055622*** (0.0312068)
Pseudo R-squared	0.0260	0.0036	0.0003
Sample Size	8258	8258	8258

Note: *Numbers in parentheses are standard errors*

****p<0.001, **p<0.01, *p<0.05*

Dependent Variable: Variety of Supermarkets (0 to 3) as described in 4.2

Table 11 - Regressions with Variety of Supermarkets as the dependent variable

Even for model 1, where all variables are significant, the much lower (pseudo) R-squared stands out when comparing the variety with the density measurement. While the variables explain, depending on the model, at least 20% of the variability of the density measurement, they only account for 2.6% of the variability of the variety measurement. This may be seen as a hint that there are no strong associations between socioeconomic characteristics and supermarket variability.

Nevertheless, all variables in model 1 have a significant impact on Supermarket variety. The results are in line with the results of chapter 6.2. Neighbourhoods with higher average income are associated with more supermarket variety. In addition, the higher the percentage of households below or around the social minimum, the lower the supermarket variety. Also, higher percentages of immigrants and elderly people are associated with more supermarket variability.

However, after increasing the threshold from 3 kilometres to 5 or even 8 kilometres, more and more variables become insignificant. A possible explanation can be described with an example: As long as the threshold is 3 kilometres, not only neighbourhoods where all three nearest supermarkets are from the same chain are specified as a ‘1’ but also neighbourhoods which only have one supermarket in 3 kilometres distance. When this threshold is increased to 8 kilometres, the measurement increasingly only takes the value of ‘1’ if all three nearest supermarkets are from the same chain. In this case, the model becomes insignificant.

Therefore, the results are twofold: First, there are associations between the number of supermarkets from different chains people have access to in a radius of 3 kilometres and the variables included in the model. Most noticeable households below or around the social minimum may have worse access. Second, there are no associations between supermarket variability and the variables included for larger radiuses. This means that the results measured in model 1 may have nothing to do with the variability of chains, but only with the number of supermarkets in close proximity. Thus, no final conclusion can be drawn about the relationship between supermarket variability and socioeconomic or demographic characteristics.

7. Discussion

As the results in chapter 5 & 6 are contradictory in some parts, this chapter serves to summarize the main findings. Furthermore, the results are compared to previous studies which conducted similar analyses in other countries, cities or regions. Thereby, the robustness of the results may be better understood and starting points for further research concentrating on specific socioeconomic indicators can be identified. Furthermore, a number of drawbacks and limitations to the study are presented. In this context, any generalisations of the results should only be done with special care.

7.1 Summary of the Main Findings

To summarize the results, Table 12 shows the results for the hypotheses which were set up previously for the different approaches. As no conclusions could be drawn from the variety measurement, the results are not included here. Thus, hypothesis 4 is rejected and will not be further interpreted. For each hypothesis, all unambiguous results throughout the models and measurements are discussed and connected with the hypothesis from chapter 3. In addition, discrepancies between the results of the models are examined and possible reasons are given.

	Graphical Analysis	Proximity Measurement	Density measurement
Hypothesis 1: • *Worse access to food stores in low-income areas* • *Lower house prices result in better access to supermarkets*	Results depend on urbanicity: low-income neighbourhoods are closer to supermarkets in urban areas, but further in Urban Clusters & Rural Areas. Worse access for low-income neighbourhoods is especially pronounced in Urban Clusters. The effect of house prices was not examined.	Higher-income neighbourhoods are significantly associated with shorter distances to the next supermarket. Households below/ around the social minimum are *further from the next supermarket*. However: Results are not always significant after incorporation of spatial effects. Lower house prices are significantly associated with smaller distances to the next supermarket.	Higher income neighbourhoods are significantly associated with more supermarkets in all tested radiuses. Households below/ around the social minimum are significantly related to *more supermarkets* in all tested radiuses. Lower house prices are significantly associated with more supermarkets in all tested radiuses.
Hypothesis 2: • *Migrants have worse access to food stores compared to natives*	High-share immigrant neighbourhoods are closer to supermarkets than the average. Exception: In rural areas, a number of high-share immigrant neighbourhoods are located further from a supermarket than the average.	Both higher shares of western- and non-western-immigrants in neighbourhoods are significantly associated with smaller distances to supermarkets throughout all models.	Both higher shares of western- and non-western-immigrants in neighbourhoods, except for non-western-immigrants in one model, are significantly associated with more supermarkets.

Hypothesis 3: • *Elderly people live, on average, closer to supermarkets*	High-share elderly neighbourhoods are located closer to supermarkets than the average. Results are especially pronounced in Urban Clusters.	Higher shares of elderly in a neighbourhood are significantly associated with smaller distances to supermarkets throughout all models.	In a radius of 1 km, higher shares of elderly in a neighbourhood are significantly correlated with a higher number of supermarkets. However, for wider radiuses of 3 and 5 kilometres higher shares of elderly people in a neighbourhood are significantly related to a lower number of supermarkets.

Table 12 - Summary of the results of Chapter 5 & 6

In the graphical, as well as in both statistical analysis, higher income is in most cases connected with better access to supermarkets. Also, lower house prices improved access to supermarkets in all cases. However, due to varying findings, it remains unclear whether households below or around the social minimum have worse access to supermarkets. When using the proximity measurements these households have worse access, while the use of the density measurement yields opposing outcomes. One explanation is that the incorporation of the income variable already captures the effect of households below or around the social minimum. As seen in chapter 4.4.3, the variables are relatively strongly correlated. All in all, *hypothesis 1* can be accepted. Areas with lower income have worse access to supermarkets, while lower house prices improve access to supermarkets. In addition, the supermarket access seems especially bad for low-income neighbourhoods in less densely populated neighbourhoods.

Hypothesis 2, which suggests that migrants have worse access to supermarkets, has to be rejected. On the contrary, in most models areas with a higher percentage of western- or non-western-immigrants are associated with better access to supermarkets. A possible reason may be the self-selection of migrants into these areas, however, no finite explanation can be given. Some exceptions are high-share migrant neighbourhoods in rural areas which are especially far from the next supermarket.

Hypothesis 3, which claims that elderly people live closer to supermarkets than the average, can be accepted. The graphical analysis, as well as the statistical analysis with the proximity measurement, verify without exceptions better access to supermarkets for neighbourhoods with higher shares of people above 65. However, the analysis yields opposing results when using the density measurement: First, for a radius of 1 km higher shares of elderly people are correlated with *more* supermarkets. Second, for radiuses of 3 and 5 km higher shares of elderly people are associated with *fewer* supermarkets. This may be interpreted as a further confirmation of the hypothesis that elderly locate themselves close to supermarkets to keep on living an independent life. For this purpose, most likely only a supermarket in very close proximity (<1km) makes sense. Travelling longer distances may pose severe challenges for many elderly and after a certain threshold the distance does not matter for the location decision: For example, it may be irrelevant if the next supermarket is 3 or 5 kilometres away since to reach both destinations many elderly need a car.

7.2 Comparison of the Findings to Previous Studies

To classify the results into the existing literature, previous results from other countries or cities are compared to the outcomes of the present analysis. All known studies which deal with whole countries are presented and compared. Furthermore, a selection of studies analysing European cities which were already described in chapter 2 or worldwide studies which stand out due to similar study designs are contrasted.

In an extensive report to the US Congress, Ver Ploeg et al. (2009), among other things, examined the interaction of neighbourhood characteristics and food access in the USA. The results were separated for different urbanicities and varied a lot. The data they used was more detailed and more variables were included. Instead of adding income averages and percentages of migrants, the authors used indicators for racial segregation and income inequality as predictors and found associations with limited access to food stores. One reason while racial segregation was found to be correlated with limited, may be explained by differences between American and European cities: Historically, cities in the US are often described as far more racially segregated than European cities (Andersson et al., 2018). In addition, while the present study includes the percentage of people 65 and older as one predictor, the American report used an indicator which combined people 65 and older *who also* suffer under poverty.

Another nationwide study, already presented earlier, analysed differences in neighbourhood accessibility to health-related resources such as supermarkets for Sweden

(Kawakami et al., 2010). However, the approach differs significantly from the approach used here: Neighbourhoods were classified into three different deprivation categories ranging from low to high based on income, unemployment, education level and percentages of social welfare recipients. Supermarkets were more prevalent in high-deprivation neighbourhoods.

The previously described study by Helbich et al. (2017), calculated the correlations between the predictors *Share of Dutch people* & *Housing values* and *measurements for food access in Amsterdam*. While, confirming the majority of findings from this study, the authors did not find indications that neighbourhoods with a larger share of minorities have worse spatial access to healthy food. In addition, higher housing values did not improve supermarket access. A general conclusion, however, proves difficult: the authors did not conduct a regression, neighbourhood income was not included and most of their analysis is explanatory.

A similar approach to the current analysis has been conducted by researches in Copenhagen (Svatisalee, et al., 2010). In a regression, the authors included information about immigration status, age as well as income. However, in contrast to this study, no association was found. A possible reason may be the much smaller sample size since only one city (Copenhagen) was included in the sample.

A study set in New York and Maryland used neighbourhood income and migration status, however, differs in the study design (Moore & Roux, 2006). The authors found fewer supermarkets for predominately minority neighbourhoods and low-income areas, therefore, partly contradicting but also confirming the results of this study.

Lastly, a study set in British Columbia (Canada) made use of density measures as well as proximity measures and conducted several regressions (Black et al., 2011). While information about household income, minority residents and population density was also included, the study added data about single detached houses, public transit and highway exits. Thus, the study adjusts the results for zoning and sitting variables. Before this adjustment was done, higher minority percentages have either no effect or improve the supermarket accessibility. However, after these variables were added, minority composition was associated with significantly worse supermarket access or still had no effect. A higher household income was always correlated with worse supermarket access. All in all, the results are in stark contrast to the results at hand. Three possible reasons for this can be identified: First, the inclusion of variables about zoning restrictions significantly altered the results. The opposing result that higher income is associated with worse supermarket was, however, independent from whether these variables are included. Second, the study area in Canada differs in terms of the spatial distribution of supermarkets from European regions, as reviews about food deserts mention recurrently (for example Black et al., 2013). Third, while this study includes all urban and rural areas in the Netherlands, the study by Black et al. (2011) only examined one urban area in Canada. Differences between various areas in the Netherlands, especially urban and rural areas, are possible and were not fully investigated.

All in all, the available studies differ widely in their results. No other study made use of the same study design while also analysing a whole country. In addition, any comparison is extremely difficult due to the differences in study design, study location and use of predictor variables. Oftentimes the available data shapes the study design significantly.

7.3 Drawbacks and Limitations of the Analysis

In general, the results of this study should only be interpreted with care. Several limitations prevent drawing overall conclusions for the Dutch population. These limitations are also the reason why no policy advice is given. Further research is needed to understand the true relationship between food access and socioeconomic as well as demographic factors in the Netherlands as more thoroughly explained in the next chapter.

First, the study uses Euclidean distances instead of street network distances for the calculated proximity, density and variety measurements. For instance, Oliver et al (2007) show that street network distances represent actual distances more precisely. Buffers and distances based on Euclidean distances tend to overestimate food store availability and miss mobility restrictions such as highways or rivers (Helbich et al., 2017). The measurements provided by CBS, however, use street network distances. Since the results of the CBS measurements and the calculated measurements are mostly similar, the effect does not appear to be significant in this case and may be neglected.

Second, the study's goal is to examine the relationship between food access and neighbourhood characteristics. Since only supermarkets close to people's home are included in the analysis, important food sources may be missed. People may buy healthy and affordable food at farmers markets or shop near their work or children's school (Bitler & Haider, 2011). This may lead to an underestimation of food access for certain groups or areas. Nonetheless, incorporating these food sources is extremely difficult, data-intensive and, therefore, seldom done in previous research (Beaulac et al., 2009).

Third, the neighbourhoods classifications vary in size and were constructed by the municipalities themselves. Little is known about possible systematic variations among cities or urban, suburban and rural areas (Black & Macinko, 2008). These scales can affect the associations between supermarkets and socioeconomic measures (Barnes et al., 2016). Since the aggregation biases, scale and zoning effects may alter the results, different administrative areas should be used to test the results for robustness. However, no alternative administrative areas were available for this research.

Fourth, the study makes use of a cross-sectional design and is unable to establish true causal relationships between risk factors and outcomes (Black & Macinko, 2008). Thus, the study cannot determine whether neighbourhoods with certain characteristics lead to worse supermarket access, or whether individuals with certain characteristics based on their own preferences sort themselves into neighbourhoods that provide desired services.

Fifth, previous results suggest that the inclusion of variables about the neighbourhood design may change the results significantly (Black et al., 2011). In addition, other omitted variables used in previous studies may alter the results. These variables include the extent of income inequality, racial segregation and housing vacancies.

Sixth, some analyses that were conducted are based on arbitrarily chosen numbers. For example, the buffer sizes for the density and variety measurement are based on previous literature but are still somewhat arbitrary parameters. While much importance was attached to obtaining robust results for these measurements, biases cannot be ruled out. In addition, the larger the neighbourhoods, the less accurate are the calculated densities and proximities for parts of the population. For instance, for a rural area with a size of 10 square kilometres, the density measure for 1 kilometre does not even include the whole neighbourhood. Therefore, even supermarkets which are located in the neighbourhood may be missed during the analysis.

Seventh, in chapter 4.4 a number of data issues are discussed which all may have affected the results of the study. Especially the missing data for the income variables for all neighbourhoods with a population smaller than 100 people may have altered the results. These areas are probably mostly rural, and the systematic exclusion has to be kept in mind when interpreting the results.

Eighth, while spatial regression models were partly incorporated, these models did not take care of the spatial autocorrelation among the error term since the p-value of Moran's I remained significant. In addition, no spatial models were applied when Poisson regression was used. The outcomes may change significantly if spatial models would have been adopted for all models.

7.4 Starting Points for Further Research

Due to the mostly significant results from this study, and the scarce literature in the Netherlands, further research should test the robustness for smaller areas. Researchers from Utrecht and Dresden already conducted an analysis for Amsterdam (Helbich et al., 2017). Similar analyses for the remaining Dutch regions can shed light on the true relationship between neighbourhood characteristics and food accessibility in the Netherlands. Furthermore, these studies should incorporate smaller neighbourhoods which were not created by municipalities. The use of neighbourhoods which are based on grids may prevent any administration biases.

One of the main reasons for researching this relationship is the assumption that better food accessibility results in a better diet and, thus, better health. Nevertheless, it is unclear whether this relationship truly exists in the Netherland since no study so far has examined it. For the rest of Europe, many studies do not identify a relationship or only a small one (see chapter 2.1). Therefore, studies investigating this relationship in the Netherlands are needed.

Previous studies in the US identify a larger amount of non-chain and smaller grocery stores in low-income areas (Ver Ploeg et al., 2009). Possibly, this could be a reason why the present study identifies worse access for low-income areas. It may be the case that there are more smaller, non-chain stores in low-income areas. Future studies may incorporate these kinds of stores and test for their prevalence in low-income neighbourhoods.

Lastly, future research should correct the limitations discussed in the previous chapter. Most of these limitations can be eliminated by the use of more detailed data on a lower scale. The limitation that no causal relationship can be established may be overcome by a longitudinal study or by doing surveys to understand why people choose certain neighbourhoods over others.

In conclusion, future studies should investigate smaller study areas, the relationship between food access and health in the Netherlands and include other possible food sources.

References

Andersson, E. K., Lyngstad, T. H., & Sleutjes, B. (2018, March). Comparing Patterns of Segregation in North-Western Europe: A Multiscalar Approach. *European Journal of Population*, pp. 151-168.

Barnes, T. L., Colabianchi, N., Hibbert, J. D., Porter, D. E., Lawson, A. B., & Liese, A. D. (2016). Scale effects in food environment research: Implications from assessing socioeconomic dimensions of supermarket accessibility in an eight-county region of South Carolina. *Applied Geography Vol 68*, pp. 20-27.

Beaulac, J., Kristansson, E., & Cummins, S. (2009). A systematic review of Food Deserts, 1966-2007. *Preventing Chronic Diseases Vol 6 No 3*.

Bitler, M., & Haider, S. J. (2011). An Economic View of Food Deserts in the United States. *Journal of Policy Analysis and Management Vol 30 No 1*, pp. 153-176.

Black, C., Moon, G., & Baird, J. (2013). Dietary inequalities: What is the evidence for the effect of the neighbourhood food environment? *Health & Place 27*, 229-242.

Black, J. L., & Macinko, J. (2008). Neighborhoods and obesity. *Nutrition Reviews Vol 66(1)*, pp. 2-20.

Black, J. L., Carpiano, R. M., Fleming, S., & Lauster, N. (2011). Exploring the distribution of food stores in British Columbia: Associations with neighbourhood socio-demographic factors and urban form. *Health & Place 17*, pp. 961-970.

Callaghan, M., Molcho, M., Gabhainn, S. N., & Kelly, C. (2015). Food for thought: analysing the internal and external school food environment. *Health Education Vol 115 Issue 2*, 152-170.

CBS. (2018). *1 in 5 obese adults satisfied with body weight*. Retrieved from https://www.cbs.nl/en-gb/news/2018/37/1-in-5-obese-adults-satisfied-with-body-weight

Charreire, H., Casey, R., Salze, P. S., Chaix, B., Banos, A., Badariotti, D., . . . Oppert, J.-M. (2010). Measuring the food environment using geographical information systems: a methodlogical review. *Public Healht Nutrition 13(11)*, pp. 1773-1785.

Clarke, G., Eyre, H., & Guy, C. (2002). Deriving Indicators of Access to Food Retail Provision in British Cities: Studies of Cardiff, Leeds and Bradford. *Urban Studies Vol 39 No 11*, 2041-2060.

Cliff, A. D., & Ord, J. K. (1995). 1973: Spatial autocorrelation. *Progress in Human Geography Vol 19: 245*.

Cummins, S., & Macintyre, S. (2002). Food deserts - evidence and assumption in health policy making. *BMJ 325(7361)*, 436-438.

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series 949*.

Das, S. P. (2007). *Microeconomics for Business.* New Delhi: Sage Publications India Pvt Ltd.

Drewnoski, A., Moudon, A. V., Jiao, J., Aggarwal, A., Charreire, H., & Chaix, B. (2014). Food environment and socioeconomic status influence obesity rates in Seattle and in Paris. *International Journal of Obesity 38*, 306-314.

Elhorst, J. P. (2014). Spatial Panel Models. *Handbook of Regional Science*, pp. 1637-1652.

Gittersohn, J., & Sharma, S. (2009). Physical, consumer, and social aspects of measuring the food environment among diverse low-income populations. *American Journal of Preventive Medicine*, pp. 161-165.

Glanz, K., Sallis, J. F., Saelens, B. E., & Frank, L. D. (2005). Healthy Nutrtion Environments: Concepts and Measures. *American Journal of Health Promotion Vol 19 No 5*, 330-333.

Helbich, M., Schadenberg, B., Hagenauer, J., & Poelman, M. (2017). Food deserts? Healthy food access in Amsterdam. *Applied Geography Vol 83*, 1-12.

Hoffman, J. P. (2016). *Regression models for categorical, count, and related variables: An applied approach.* EBSCO Academic Collection.

Ishikawa, M., Yokoyama, T., Nakaya, T., Fukuda, Y., Takemi, Y., Kusama, K., . . . Murayama, N. (2016). Food Accessibility and Perceptions of Shopping Difficulty Among Elderly People Living Alone in Japan. *J Nutr Health Aging Volume 20, Number 9*, pp. 904-911.

Jakobsen, T. G., & Mehmetoglu, M. (2016). *Applied Statistics Using Stata: A Guide for the Social Sciences.* SAGE Publications.

Jones, K., & Simmons, J. (1987). *Location, location, location: Analysing the retail environment.* Toronto: Methuen.

Kawakami, N., Winkleby, M., Skog, L., Szulkin, R., & Sundquist, K. (2010). Differences in neighborhood accessibility to healht-related resources: A nationwide comparsion between deprived and affluent neighborhoods in Sweden. *Health & Place 17*, 132-139.

Krizan, F., Bilkova, K., Kita, P., & Hornak, M. (2015). Potential food deserts and food oases in a post-community city: Access, quality, variability and price of food in Bratislava-Petrzalka. *Applied Geography Vol 62*, 8-18.

Lamb, K. E., Ogilvie, D., Ferguson, N. S., Murray, J., Wang, Y., & Ellaway, A. (2012). Sociospatial distribution of access to facilities for moderate and vigorous intensity physical activity in Scotland by different modes of transport. *International Journal of Behavioral Nutrition and Physical Activity 9:55*.

Lamb, K. E., Thornton, L. E., Cerin, E., & Ball, K. (2015). Statistical Approaches Used to Assess the Equity of Access to Food Outlets: A Systematic Review. *Public Health Vol 2 Issue 3*, 358-401.

Layte, R., Harrington, J., Sexton, E., J., P. I., Cullinan, J., & Lyons, S. (2011). Irish exceptionalism? Local food environments and dietary quality. *J Epidemiol Community Health Vol 65 No 10*, 881-888.

LeSage, J. P. (2010). *Econometric Toolbox*. Retrieved from http://www.spatial-econometrics.com/

Macdonald, L., Ellaway, A., Ball, K., & Macintyre. (2011). Is proximity to a food retail store associated with diet and BMI in Glasgow, Scotland? *BMC Public Health No 11:464*.

Macintyre, S., Macdonald, L., & Ellaway, A. (2008). Do poorer people have poorer access to local resources and facilities? The distribution of local resources by area deprivation in Glasgow, Scotland. *Social Science & Medicine 67*, 900-914.

Mantovani, R., Daft, L., Macaluso, T. F., & Welsh, J. H. (1997). Authorized Food Retailers' Characteristics and Access Study. *US Department of Agriculture: Alexandria VA*.

McCann, P. (2013). *Modern Urban And Regional Economics* (Second Edition ed.). Oxford University Press.

Molaodi, O. R., Leyland, A. H., Ellaway, A., Kearns, A., & Harding, S. H. (2012). Neighbourhood food and physical activity environments in England, UK: does ethnic density matter? *International Jounral of Beharioral Nutrition and Physical Activity 9:75*.

Moore, L. V., & Roux, A. V. (2006). Associations of Neighborhood Characteristics With the Location and Type of Food Stores. *American Journal of Public Health Vol 96 No 2*, pp. 325-331.

Moran, P. A. (1950). A test for the serial independence of residuals. *Biometrika Vol 37 No 1-2*, 178-181.

OECD. (2017). *OECD Social and Welfare Statistics: Income distribution*. Retrieved from https://data.oecd.org/inequality/income-inequality.htm

Oliver, L. N., Schuurman, N., & Hall, A. W. (2007). Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *Int J Health Geogr 6:41*.

Pearson, T., Russel, J., Campbell, M. J., & Barker, M. E. (2005). Do 'food deserts' influence fruit and vegetable consumption?—a cross-sectional study. *Appetite Vol 45 No 2*, 195-197.

Plus. (2019, April 24). Retrieved from https://www.plus.nl/info-over-plus/organisatie

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods Vol 7 No 2*, 147-177.

Shaw, H. (2012). Access to healthy food in Nantes, France. *British Food Journal Vol 114 Issue 2*, 224-238.

Stata. (2019). *Stepwise estimation.* Retrieved from https://www.stata.com/manuals13/rstepwise.pdf

Statistics Netherlands. (2015). *Transport and mobility.*

Statline, C. (2018, December 24). *Life style and (preventive) health examination; personal characteristics*. Retrieved from https://opendata.cbs.nl/#/CBS/en/dataset/83021ENG/table?dl=11D79

Stelder, D. (2012). Spatial monopoly of multi-establishment firms: An empirical study for supermarkets in the Netherlands. *Papers in Regional Science Vol 91 No 1*, 181-192.

Svatisalee, C. M., Nordahl, H., Glümer, C., Holstein, B. E., Power, L. M., & Due, P. (2010). Supermarket and fast-food outlet exposure in Copenhagen: associations with socio-economic and demograhpic characteristics. *Public Health Nutritions Vol 14 No 9*, 1618-1626.

Toelichting Wijk- en Buurtkaart 2015, 2016 en 2017. (2017, August). Retrieved from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische%20data/wijk-en-buurtkaart-2015

Ver Ploeg, M., Farrigan, T., Breneman, V., & Hamrick, K. S. (2009). *Access to Affordable and Nutritious Food—Measuring and Understanding Food Deserts and Their Consequences: Report to Congress.* Economic Research Service/ United States Department of Agriculture.

Walker, R. E., Keane, C. R., & Burke, J. G. (2010). Disparities and access to healthy food in the United States: A review of food deserts literature. *Health & Place 16*, 876-884.

Walsh, P. (2013). *Mathworks*. Retrieved from Nearest Neighbour Spatial Weights Matrix: https://www.mathworks.com/matlabcentral/fileexchange/36601-nearest-neighbor-spatial-weights-matrix

White, M., Bunting, J., Williams, L., Raybould, S., Adamson, A., & Mathers, J. (2004). Do 'food deserts' exist? A multi-level, geographical analysis of the relationship between retail food access, socio-economic position and dietary intake. *Food Standards Agency*.

Williamson, S., Mcgregor-Shenton, M., Brumble, B., Wright, B., & C., P. (2017). Deprivation and healthy food access, cost and availability: a cross-sectional study. *Journal of Human Nutrition and Dietetics*, 791-799.

Wrigley, N., Warm, D., & Margetts, B. (2003). Deprivation, diet, and food-retail access: findings from the Leeds "food deserts" study. *Environment and Planning A Vol 36*, 151-188.

Xue, B., Oldfield, C. J., Dunker, A. K., & Uversky, V. N. (2009). CDF it all: Consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. *FEBS Letters Volume 583 Issue 9*.

Zenk, S. N., Schulz, A. J., Israel, B. A., Sherman, A. J., Shuming, B., & Wilson, M. L. (2005). Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit. *American Journal of Public Health*, pp. 660-667.

8. Appendix

8.1 Appendix 1 – Adjustments to the Supermarket Dataset & CBS Dataset

8.1.1 Supermarket Dataset

The names of the supermarkets were scanned for various chain names as well as different spellings to identify all chain-supermarkets. The terms searched for are: *Plus, plus, PLUS, Spar, spar, SPAR, Coop, coop, COOP, Coöp, Cöop, Lidl, lidl, LIDL, Aldi, aldi, ALDI, Jumbo, jumbo, JUMBO, c1000, C1000, C 1000, C-1000, Dirk van, dirk van, Dirk Van, DIRK VAN, Ekoplaza, ekoplaza, EKOPLAZA, EMTÉ, EM- TÉ, EM-Té, Em-Té, em-té, EMTE, Em-Te, Em/TÉ, Emte, Emté, Emt_, Boni, boni, BONI, Deen, DEEN, Hoogvliet, Albert Heijn, AH, ALBERT HEIJN, A.H., Albert heijn, Albert Heyn, Ah, Heyn Albert, Deka, JAN LINDERS, Jan Linders, Nettorama, NETTORAMA, POIESZ, Poiesz, Vomar, Bio* or bio

All bio-stores (including Ekoplaza) and all remaining stores were excluded from the analysis.

All store names containing *to go, togo, TOGO, TO GO* or *To Go* were excluded from the analysis.

All stores with no available coordinates were excluded from the analysis

8.1.2 CBS Dataset

All areas marked as containing only water were dropped from the analysis such as the Ijsselmeer (WATER=’JA’). In addition, all negative observations were changed out to ‘missing’ and later excluded if needed in a regression.

8.2 Appendix 2 – Average Distance to the Next Closest Supermarket

The given 1.1-kilometre average distance to the next supermarket for the Netherlands is the average distance of all neighbourhoods, while the UK study uses average household distance. However, in the Netherlands after weighting each neighbourhood’s distance depending on the number of residents living there, the distance decreases even further to only 847 meters.

x

8.3 Appendix 3 – Graphs Using the Calculated Distance from the Neighbourhood Centroid to closest Chain-Supermarket

Figure 15 - Distance to Chain-Supermarket by Urbanicity

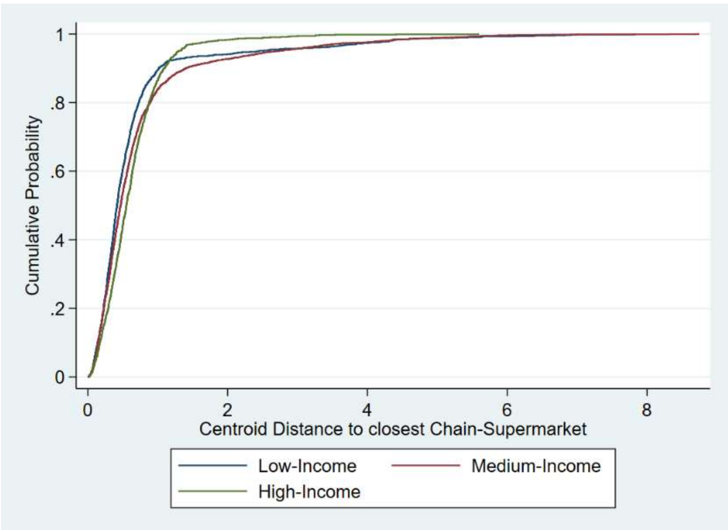

Figure 16 - Distance to closest Chain-Supermarket by Income group in Urban Areas

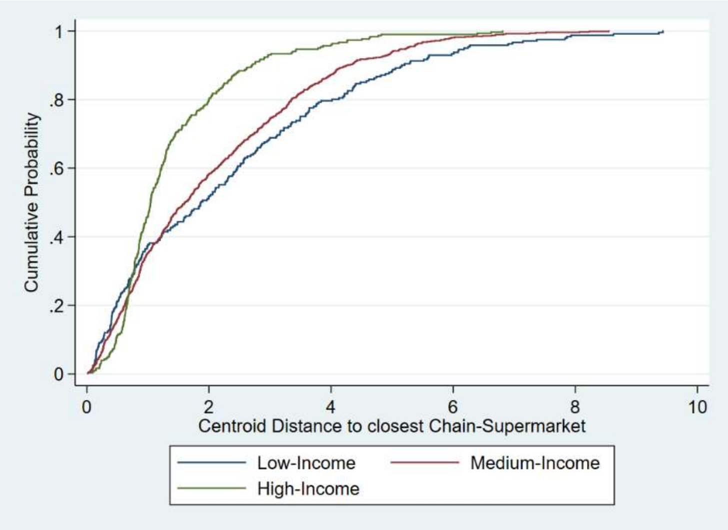

Figure 17 - Distance to closest Chain-Supermarket by Income group in Urban Clusters

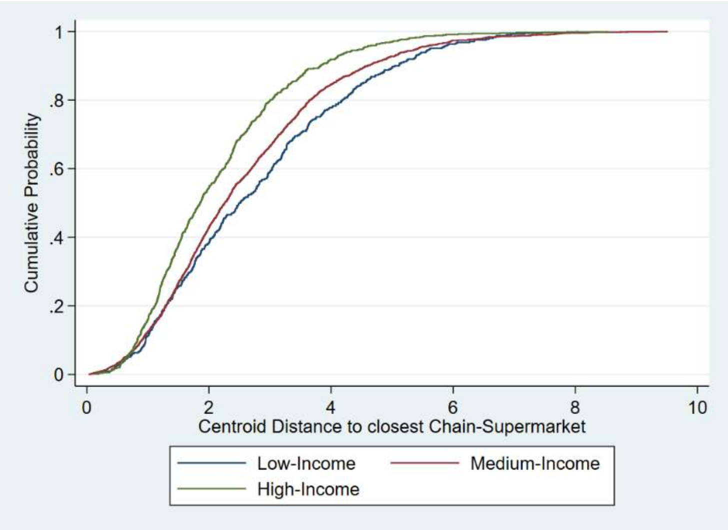

Figure 18 - Distance to closest Chain-Supermarket by Income group in Rural Areas

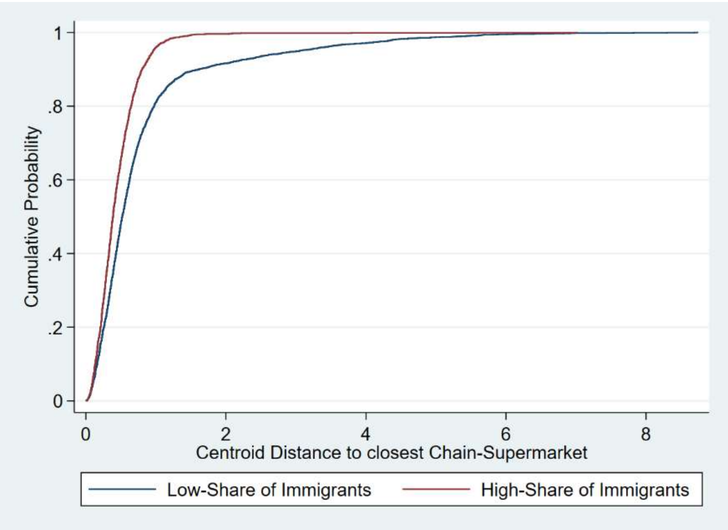

Figure 19 - Distance to closest Chain-Supermarket by Migration Status in Urban Areas

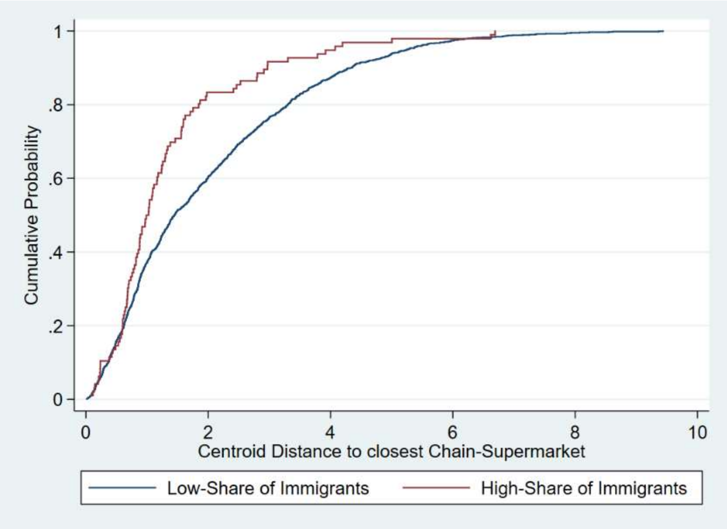

Figure 20 - Distance to closest Chain-Supermarket by Migration Status in Urban Clusters

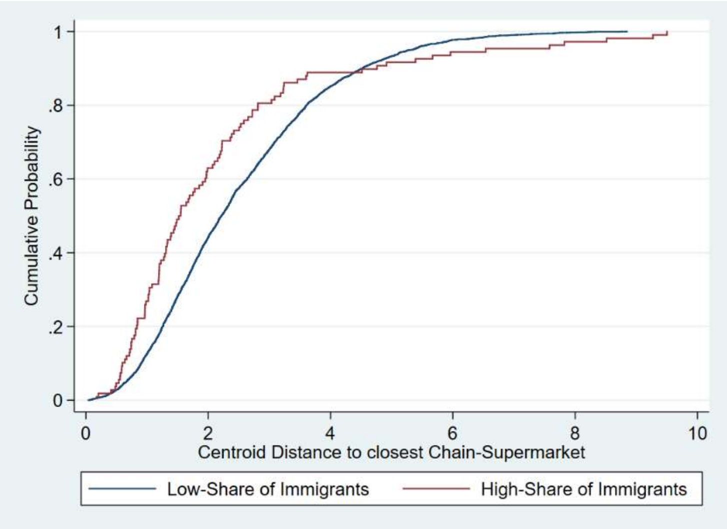

Figure 21 - Distance to closest Chain-Supermarket by Migration Status in Rural Areas

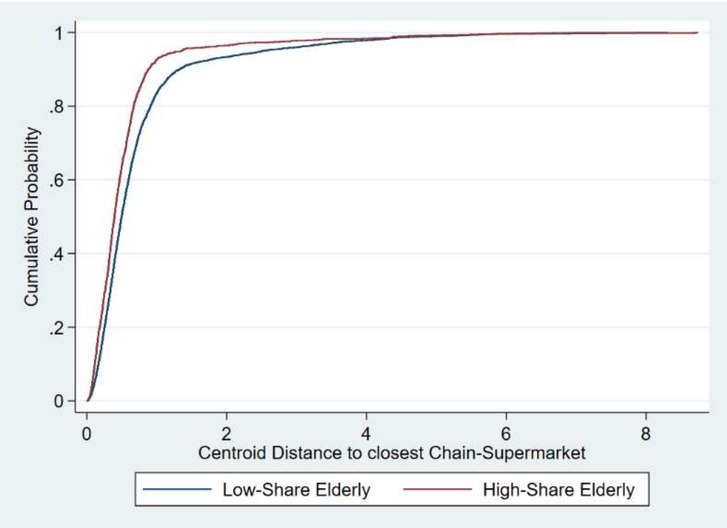

Figure 22 - Distance to closest Chain-Supermarket by Elderly Share in Urban Areas

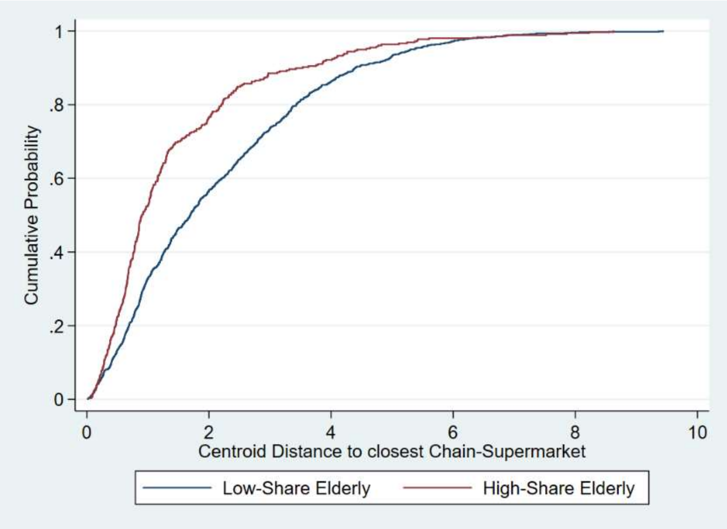

Figure 23 - Distance to closest Chain-Supermarket by Elderly Share in Urban Clusters

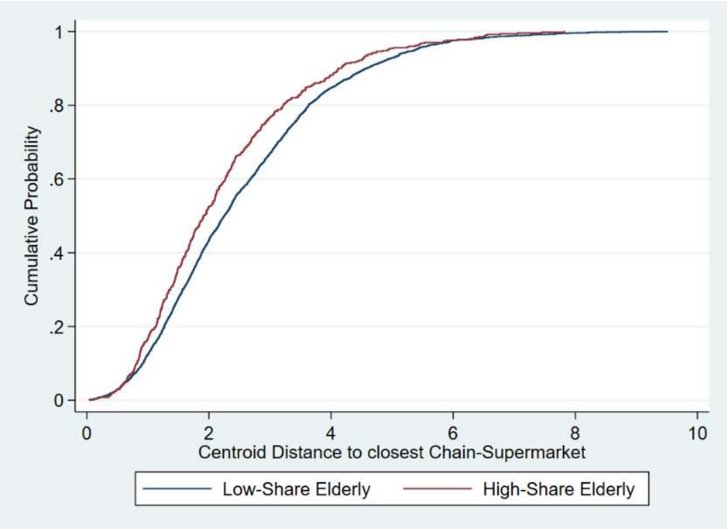

Figure 24 - Distance to closest Chain-Supermarket by Elderly Share in Rural Areas

8.4 Appendix 4 – Spatial Autocorrelation between Attributes

Variable	**Moran's Index**	**z-score**	**p-value**
Population	0,206703	66,530418	0,000000
Percentage of people 0 to 15 years	0,084763	27,277093	0,000000
Percentage of people 65 and older	0,085110	27,397625	0,000000

Population density per km2	0,517193	166,326238	0,000000
Percentage of people with a western migration background	0,291195	93,762958	0,000000
Percentage of people with a non-western migration background	0,415597	133,719655	0,000000
Average house value	0,133950	43,095427	0,000000
Average Income per inhabitant	0,099807	32,112349	0,000000
Percentage of households with the lowest income	0,184770	59,422586	0,000000
Percentage of households below or around social minimum	0,224531	72,219885	0,000000
Large Supermarket average distance in km	0,260014	83,621062	0,000000
large supermarket average number within 1 km	0,437302	140,661466	0,000000
large supermarket average number within 3 km	0,766114	246,500397	0,000000
large supermarket average number within 5 km	0,910698	292,993820	0,000000
Area	0,092368	29,810768	0,000000

Table 13 - Spatial Autocorrelation between Attributes

The null Hypothesis of no spatial autocorrelation cannot be rejected for any of the variables. All variables are positively spatial autocorrelated. Since all Z-scores are positive the spatial distribution of high values and low values is more spatially clustered than if the process would be random. In this case the high spatial autocorrelation for the variables makes sense: For example, neighbourhoods with a high population density are mostly located in inner cities, surrounded by other inner-city neighbourhoods with high population densities. The same is valid for all other variables: There is no reasonable explanation why neighbourhood characteristics should change drastically between adjacent neighbourhoods since the borders are more or less arbitrarily defined by each municipality.