Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Analyzing the influence of housing, urbanization and regional indicators on mental health, physical health and dementia

Nori van den Biggelaar

MSc 06/2021-002

Analyzing the influence of housing, urbanization and regional indicators on mental health, physical health and dementia

By Nori van den Biggelaar
SNR: 1279708

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Econometrics and Operations Research

Tilburg School of Economics and Management
Tilburg University

Supervised by
Tilburg University: dr. Tunga Kantarcı
Republiq: ir. Peter Korst MRE

June 11, 2021

Abstract

Housing associations own approximately 2.4 million houses in the Netherlands. The first housing association in the Netherlands was established in the late 1800's in order to improve the health of the working class. Lately, the focus of housing associations has shifted from the health of the people living in the houses, to the financial state and sustainability of the houses. In order to advise housing associations on how to shift the focus back to improving the health of the people living in social dwellings, this paper aims to establish relationships between housing and indicators for urbanization and provinces, and five different health outcomes: psychiatric issues, intellectual disability, somatic issues, mobility issues and dementia. We use a linear regression model, estimated by Ordinary Least Squares (OLS). Analyzing correlations, we find that a large fraction of small and old houses, which we could define as bad quality houses, leads to a higher probability of having mental and physical health problems. Furthermore, the probability of having mental health issues and dementia is associated with higher levels of urbanization, while the probability of having physical health issues is associated with lower levels of urbanization. We find that a large fraction of social dwellings in a neighborhood is associated with mental and physical health problems, even when controlling for income, while this is not the case for dementia.

1 Introduction

In the 19th century, Europe went through an era of industrialization and urbanization. Besides all the good that this era has brought Europe, it also had its downsides. People were exploited, cities were overcrowded and so-called slumlords owned a large amount of houses. They housed people in barns and shacks for extortionate prices. Regulations on hygiene were absent, which led to wide-spread cholera-epidemics and other diseases. Overall health of the working class and their families was extremely bad (Canon Volkshuisvesting, 2015).

The urban elite in London were the first people to come up with a plan to offer good and affordable housing. In 1852, the Netherlands followed with an initiative called 'de Vereeniging ten behoeve der Arbeidersklasse te Amsterdam', which roughly translates to 'the association for the benefit of the working class in Amsterdam': the first housing association was established. (Canon Volkshuisvesting, 2015).

In the 19th century, the relationship between housing and health was not yet very evident. The Dutch Royal Institute of Engineers in 1854 plead for better housing for the working class, with enough space and good ventilation. They thought that this would lead to better health of working people, which in turn would lead to economic growth. However, only in 1901 the Housing Act was introduced. This law made it, among other things, possible for housing associations to borrow money to build houses. The main reason for this was that in order to improve the health of the working class, better quality houses had to be built. The Housing Act paid off: in 100 years around 2.4 million social dwellings were built, and nowadays there are around 300 housing associations.

Although health of the working class was one of the main reasons housing associations were established, nowadays the focus is less and less on the people living in the social dwellings. It has shifted to the financial situation and the sustainability of the houses. Back in the 19th century and the beginning of the 20th century, the houses were the means, not the goal themselves. Nowadays houses have to be renovated, they have to be made sustainable to meet the goals according to the Paris Climate Agreement, and the rent should remain low since people with lowest incomes are eligible to accommodate in the social dwellings. The focus on the well-being and health of the people living in social dwellings is moving to the background, while this should be the main focus of housing associations.

In this paper, we aim at explaining five different health outcomes: psychiatric problems, intellectual disabilities, somatic problems, mobility problems and dementia. We consider house characteristics, household characteristics, age, income and regional indicators to analyze what influences these health outcomes. The goal of this paper is to discover relationships between the explanatory variables and the health outcomes, which can be used to draw policy implications for housing associations on how to have more influence on the health of the people living in their social dwellings.

Republiq is a consultancy firm, advising parties on real-estate matters in the semi-public sector. They for example consult municipalities on facility planning, educational institutions on an integral housing plan or housing associations on how to make their houses sustainable. Republiq is a unique firm in several respects. First, they have a large amount of data available on all subjects regarding real estate (and the people living in it). Besides, they have a unique view on real estate: not only financial matters are important, but all stakeholders, for instance the people renting social dwellings, are. This means that when consulting a public institution, the focus is not only on the buildings, but also on the people who live in these buildings. This makes Republiq an important business partner for governments, municipalities, educational institutes and housing associations.

In the literature, a number of studies have analyzed the relationship between mental health

and house characteristics. Evans et al. (2003) state that people living in high-rise buildings or apartments have more mental health issues than people living in low-rise buildings or houses, which may be due to social isolation and a lack of space. Residents of single family detached homes typically have the best mental health (Evans et al., 2003). Evans et al. state that housing quality has a positive correlation with psychological well-being. People in high-rise buildings have less sense of safety, less attachment to the community and less social support (McCarthy and Saegert, 1987).

According to the literature, living alone also affects on mental health. People who experience loneliness or isolation are more likely to exhibit depressive symptoms (Smith and Victor, 2018). Loneliness is associated with psychoses, personality disorders and cognitive decline over time (Hawkley and Cacioppo, 2010). Feelings of social disconnectedness and loneliness also increase stress, anger and anxiety (Cacioppo et al., 2006).

Sundquist and Ahlen (2006) showed that people living in poorer neighborhoods in Sweden had a significantly higher risk of being hospitalized with mental health problems rather than people living in richer neighborhoods, when controlling for socioeconomic and demographic characteristics. It is also the case that the impact of urbanization is affiliated with an increase in mental health problems (Srivastava, 2009).

Research has shown that poor quality housing has an effect on physical health too. Fuller et al. (1993) states that for example that high levels of household crowding, which can be the result of living in a small house, can lead to stress which can in turn lead to physical illnesses. Shaw (2004) shows that housing affects health in a lot of minor ways, and that it in total is one of the key determinants of physical health.

Research in Lebanon shows that there is a significant positive association between chronic illness and housing conditions (Habib et al., 2008). Next to this, research from the United States indicates that every additional poor quality housing characteristic leads to poorer health status, higher likelihood of hospitalization and a higher medical utiliziation (Boch et al., 2020). Boch et al. state that 'home is where the health is'.

Smith and Victor (2018) find that people who are living alone, lonely or in isolation are the people with the highest proportion of multi-morbidity (having two or more chronic conditions). Furthermore, the groups of people that reported loneliness were also the groups of people with the highest proportion of two or more problems with activities of daily living.

Income also seems to have a relationship with physical health. People in higher income groups are experiencing better physical health (Johnson and Krueger, 2005). It is the case that the more advantaged people are, the better level of health they experience (Adler and Ostrove, 1999).

Literature on dementia in relationship to housing and urbanization is firstly not widely available, and secondly it is contradictory. For instance, Robbins et al. (2019) find that on the one hand, rates of dementia among people living in urban regions are lower than people living in rural regions. However, they also find that greater exposure to pollutants, both noise and air, which is common in urban areas, may negatively affect cognition. This could lead to a cognitive disease such as dementia.

This research contributes to the existing literature in several respects. First, we have a rich source of information, which makes it possible to exploit much variation in the health outcomes. The explanatory variables in this paper are extensive.

Next to this, in this paper we use objective health data instead of subjective health data. The health outcomes are based on reported health data in hospitals and with general practitioners, instead of being based on survey data. This fact, together with the fact that this health data covers the entire Netherlands makes this paper unique.

Furthermore, in this paper we look at the relationship between health and urbanization in

the Netherlands. This relationship is not yet much researched in the Netherlands, which makes this paper a good contribution to the existing literature.

Lastly, in this paper, dementia among people that live independently is one of the health outcomes. Therefore, this paper provides insight in how housing characteristics and urbanization are related to demented people that live at home. As mentioned before, there is not much literature available on dementia in relationship to housing and urbanization. This makes this paper an interesting contribution to existing literature.

We find that bad quality housing is associated with a higher probability of having mental and physical health problems. Furthermore, the larger the fraction of social dwellings in a neighborhood is, the higher the probability of having mental en physical health problems is. We also find that people with dementia that do not live in care facilities live in relatively large and old houses. Lastly, we find that there is regional variation in all five health outcomes. In very urbanized areas, both the probability of having mental health problems and the probability of having dementia is high, while the probability of having physical health issues is low. Living in any region but Groningen is associated with a lower probability of having mental and physical health issues.

The remainder of this paper is organized as follows. In Section 2 we zoom in on all data and variables we use for the analysis and the way this data was manipulated. Section 3 presents descriptive statistics. In Section 4 we elaborate on the empirical approach that we use in this thesis. We display and discuss the results of the analysis in Section 5. Section 6 presents conclusions and a discussion.

2 Data

In this paper, we use house characteristics, household composition variables, age and income to predict health outcomes. We combine data from a number of institutions, which include the Central Bureau of Statistics (CBS), the Health Monitor, the Care Needs Assessment Center, Kadaster, PBL Netherlands Environmental Assessment Agency, the Ministry of Finance and In.Fact Research. The health outcomes in this paper come from In.Fact Research, which is a Dutch independent research institute. On behalf of the Dutch government, they collected data on different kinds of health outcomes for the entire Dutch population, which they called the Woonzorgwijzer. The Woonzorgwijzer data is collected in 2012, 2014, 2016 and 2018, but we only have access to the data of 2018.

The variables in the Woonzorgwijzer are the outcome variables in this paper. The variables are defined as probabilities of experiencing a particular health problem. These variables are based on data from CBS, the Health Monitor and the Care Needs Assessment Center (Centrum Indicatiestelling Zorg). The CBS data contains micro-data on all people in the Netherlands, such as their age, income, gender et cetera. The data in the Health Monitor and the Care Needs Assessment Center is health data that is reported in hospitals and at general practitioners.

We use house characteristics, variables on household composition, age and income as independent variables in this paper. The house characteristics include the construction year of the houses, the surface of the houses and the type of house for all houses in the Netherlands. It also includes an indicator for whether a house is owned by a housing association. This data is collected in 2019.

The variables on household composition are based on a PEARL-prognosis, which is a household prognosis. These numbers are from 2018. Control variables on age of people are also from the year 2018.

An important control variable in this thesis is the income variable. Past research has shown that income is an important indicator for health problems, as people with lower incomes have

less access to healthcare (Adeline and Delattre, 2017). The income data is based on personal income. This contains labor income, but also income from social insurance. CBS based this data on data from the Ministry of Finance and the Population Register of the Netherlands.

Urbanization variables are included in the paper as explanatory variables, since we expect variation in health outcomes for different urbanization levels. These variables are constructed by the CBS as well and indicate how urbanized a region is. This variable is based on the surrounding address density of a region, which is the average number of addresses per square kilometer.

We perform a number of manipulations on the data. First, we aggregated the variables we have on house-level to the neighborhood-level. The reason for this is that except for the house characteristics, all variables are available on neighborhood-level. All house characteristics from Kadaster are available on the house-level. This means that for each house in the Netherlands, with a total of about 8 million homes, we know the construction year, the house type and the area. For each neighborhood in the Netherlands, we compute the fraction of houses that have a construction year in a certain category, an area in a certain category or a certain house type. The reason we use fractions is that the dependent variables from Woonzorgwijzer are also relative to the number of people living somewhere, since those variables are probabilities.

We deleted observations that led to inconsistencies in the data. We have removed the houses with an area below 10 square meters, an area above 1000 meters, a construction year below 500 and above 2020 (since the data is from 2019). These restrictions lead to dropping 5731 of 8 million houses.

Next, we matched the aggregated house characteristics variables with the dependent variables on the neighborhood level. There are about 13,000 neighborhood in the Netherlands. However, the dependent variables from Woonzorgwijzer are only available for 9891 neighborhoods. This is because of the fact that in some neighborhoods there are only a few houses, and due to privacy reasons the health measures are not available.

The income variable is not available for 21 neighborhoods, which is also due to data privacy restrictions. The age variable is not available for 13 neighborhoods, for unknown reasons. These missing values are randomly spread across the Netherlands, meaning that deleting those missing values will not bias our results. This leads to 9857 neighborhoods which constitutes our study sample.

Outcome variables

The outcome variables are five health measures. The health data was not collected using a survey, but using data on diagnoses that official doctors (for instance general practitioners) made. In.Fact collected micro-data from CBS for all people in the Netherlands. This data consisted of for instance age, gender, address, income, household characteristics, living conditions and so on. Health data is not available for all people in the Netherlands. Using regression analysis, In.Fact Research extrapolated the available health data to all people in the Netherlands.

This extrapolation was tested. The researchers used the model they created to extrapolate the health data to predict true values for all individuals in the Netherlands. Then, the explained variation was computed, which is the correlation between the predicted values and the actual values squared. They computed this explained variation for all different kinds of health outcomes, except for the probability of having an intellectual disability. The explained variation for intellectual disabilities was estimated on basis of literature. They also did not include information on severe dementia, since the Woonzorgwijzer only concerns people living independently and people with severe dementia almost never live independently. The available explained variations are displayed in Table 1. As shown in Table 1, each independent variable

was constructed using data on multiple kinds of health issues. To perform the research in this paper, we assume that the predicted health data is correct.

Table 1: Explained variations of dependent variables

Dependent variables	Expl. variation (%)
Probability of having psychiatric issues:	
Anxiety and depression, light issues	85
Issues with self-sustainability, light issues	30
Severe issues with self-sustainability	88
Severe behavioral issues	53
Probability of having somatic/mobility issues:	
Singular somatic issues	59
Multiple som. issues, joint issues	91
Multiple som. issues, incontinence	84
Multiple som. issues, heart failure, COPD	85
Multiple som. issues, high blood pressure	86
Multiple som. issues, skin conditions	53
Physical handicap	59
Probability of having dementia:	
Mediocre dementia	71
Light dementia	75

Source: Ontwikkeling Woonzorgwijzer, Kwantificering groepen met beperkingen (2016)

To compute the probabilities of experiencing health issues, the researchers aggregated the data to a grid level, which is a square area of 100 by 100 meters. This means that they summed up the people who, according to the constructed model, experience a particular health issue, like having psychiatric issues, and divided this sum by the total amount of people in the regarding grid. For our research we aggregated the data to neighborhood level. A neighborhood is part of a municipality and consists of multiple grids. To aggregate the data, we averaged over the available grid data.

In this paper, we make a distinction between mental health problems, physical health problems and dementia. For the models on mental health outcomes, we use the probability of having psychiatric problems and the probability of having and intellectual disability as outcome variables. The probability of having psychiatric problems is the probability that someone is living independently, but experiences psychiatric problems like depression, problems with self-sustainability and/or more. The probability of having an intellectual disability is the probability that someone is living independently, but has an IQ between 50 and 85. People with an intellectual disability have limitations in certain things, like communication, self-care and social skills.

The two outcome variables in the models for physical health problems are the probability of having somatic problems and the probability of having mobility problems. Somatic problems are inner-body problems, like COPD, high blood pressure or heart failure, which make it hard to walk, carry heavy things or stand up. Mobility problems are outer-body problems, like a bad

hip. These matters lead to problems with moving around in general as well. In the fifth model in this paper, we use the outcome variable 'the probability that someone has dementia'.

All these probabilities are computed for people who live independently, that is, people living in a care facility are not included. All outcome variables have values between 0 and 1.

Explanatory variables

We have one explanatory variable for each of the following house types: detached house, semidetached house, corner house, rowhouse and apartment. It is computed as the fraction of houses in the neighborhood that is of that particular type.

The construction year of the houses is divided into four categories: 'below 1945', 'between 1945 and 1965', 'between 1965 and 1991', 'above 1991'. For each of the categories there is one variable that is the fraction of houses in the neighborhood with a construction year in that particular category. The years that determine the categories are based on Housing Acts, which are laws about quality of new-built houses, that were introduced in those years.

The next set of explanatory variables is about area of houses. For each of the categories 'below 70 square meters', 'between 70 and 110 square meters', 'between 110 and 150 square meters' and 'above 150 square meters' we computed the fraction of houses in the neighborhood that have an area in that category.

Next household composition variables are considered. For households with singles, households with two people and for households with children we computed the fraction of households in the neighborhood of that particular type.

We did the same for age. That is, for age categories 0 to 45 years, 45 to 65 years and 65+ years, we computed the fraction of people in the neighborhood of that particular age.

Next, we consider urbanization dummies as explanatory variables. Every neighborhood has a particular urbanization level. Very strongly urbanized means that the surrounding address density (SAD) is higher than 2500 addresses per square kilometer. The neighborhood is strongly urbanized if the SAD is between 1500 and 2500. The neighborhood is moderately urbanized when the SAD is between 1000 and 1500 addresses per square kilometer, slightly urbanized when the SAD is between 500 and 1000 and not urbanized when the SAD is below 500 addresses per square kilometer. Since we do not particularly expect a linear effect in urbanization levels, we constructed one dummy variable for each of the urbanization levels.

We also include a dummy for city, which is equal to 1 if the neighborhood is part of a city, and 0 otherwise.

Using the data on urbanization levels we computed the travel distance in kilometers from each of the neighborhoods to the nearest neighborhood with urbanization level 1.

Furthermore, for each of the twelve provinces in the Netherlands, we include a province dummy. This means that for example when a neighborhood is in Limburg, the dummy for Limburg is equal to 1, and all other province dummies are equal to 0.

Finally, we include an income variable and a variable for housing associations. The latter is a computed as the fraction of houses in the neighborhood that are owned by a housing association. The income variable is computed as the logarithm of the average income in the neighborhood. This income concerns the personal income, which is the sum of labor income and income from social insurance.

3 Descriptive statistics

In this section we present descriptive statistics of the dependent variables and the independent variables in this thesis.

Dependent variables

As can be seen in Figures 1 to 5, the probability of having an intellectual disability has both the highest mean and the highest maximum. The other measure of mental health, the probability of having psychiatric problems has the lowest mean and the lowest maximum. This could be because of the fact that it is easier to quantify an intellectual disability than to quantify psychiatric problems.

Except for a peak at lower values of the distributions for the probabilities of having an intellectual disability, having somatic problems and having mobility problems, all distributions except for the distribution in Figure 1 seem close to a normal distribution. The distribution of the probability of having psychiatric problems does not look like a normal distribution.

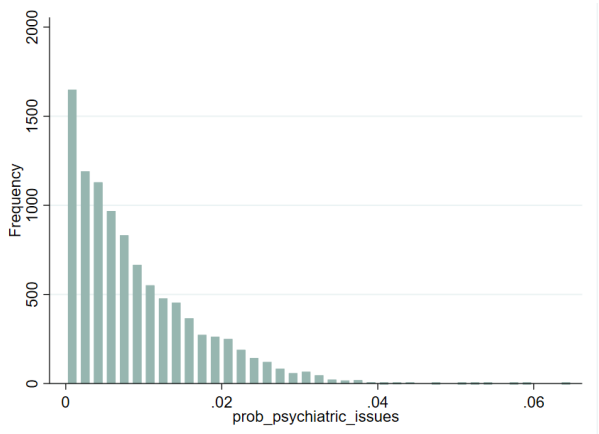

Figure 1: Distribution of prob. of having psychiatric problems

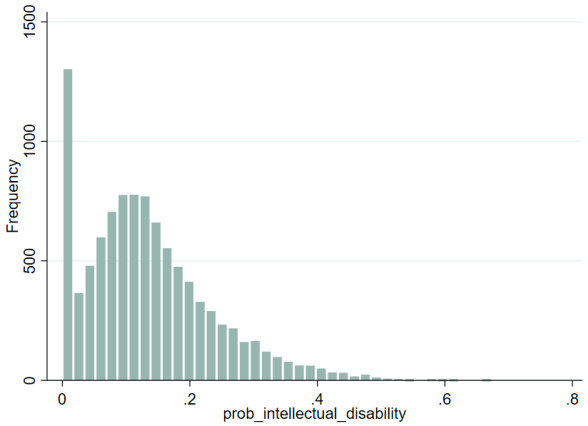

Figure 2: Distribution of prob. of having an intellectual disability

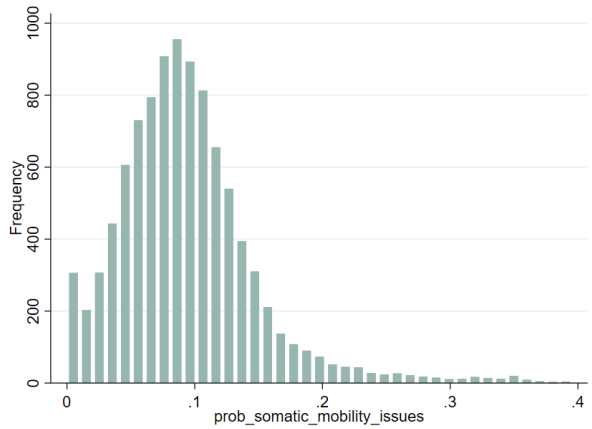

Figure 3: Distribution of probability of having somatic problems

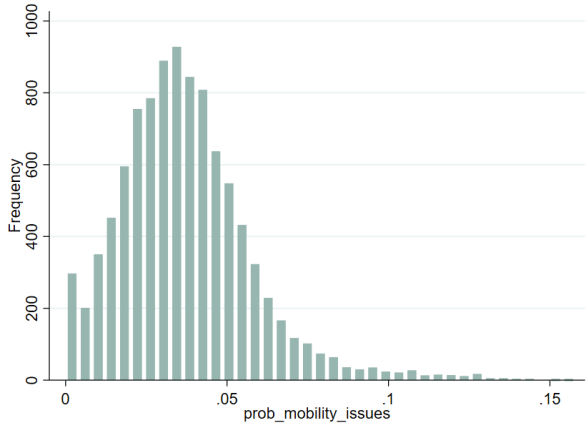

Figure 4: Distribution of probability of having mobility problems

Figure 6: Probability of having psychiatric problems

Figure 7: Probability of having an intellectual disability

Figure 8: Probability of having somatic problems

Figure 9: Probability of having mobility problems

Figure 10: Probability of having dementia

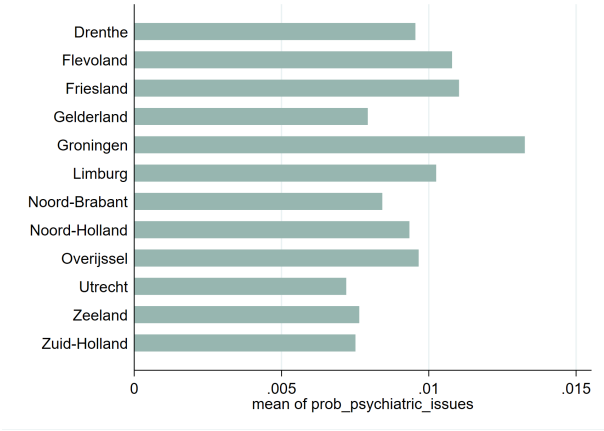

Figure 11: Mean probability of having psychiatric problems per Dutch province

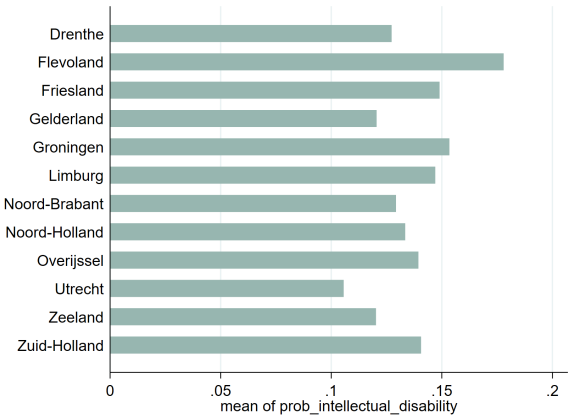

Figure 12: Mean probability of having an intellectual disability per Dutch province

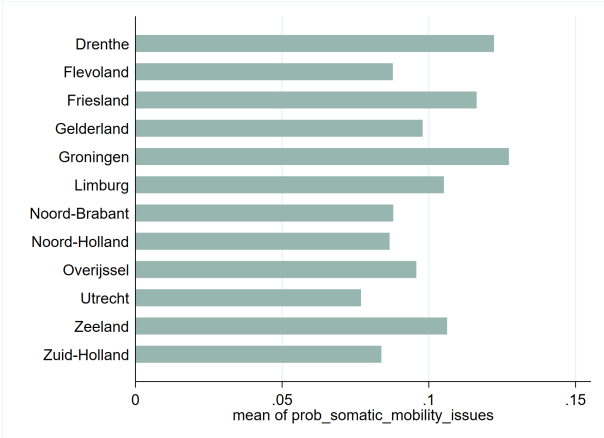

Figure 13: Mean probability of having somatic problems per Dutch province

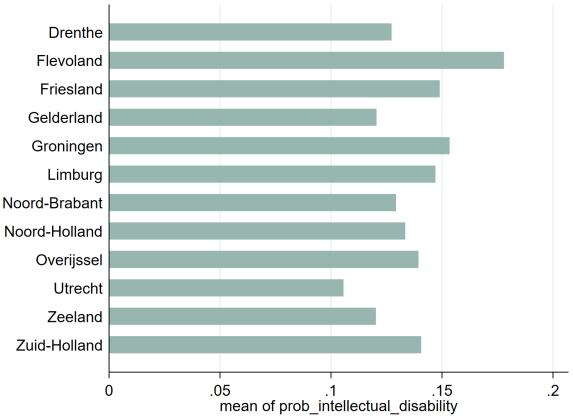

Figure 14: Mean probability of having mobility problems per Dutch province

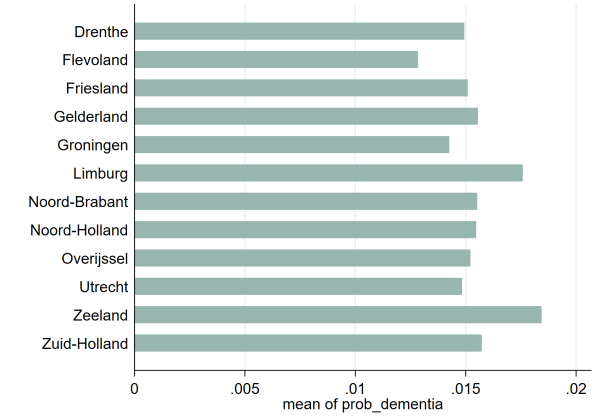

Figure 15: Mean probability of having dementia per Dutch province

Explanatory variables

Table 2: Sample statistics Explanatory variables

		Mean	Standard dev.
Fraction of houses of type	Detached	0.224	0.259
	Semi-detached	0.138	0.140
	Corner	0.171	0.098
	Row	0.362	0.240
	Apartment	0.104	0.208
Fraction of houses of construction year	Before 1945	0.214	0.279
	1945-1965	0.150	0.208
	1965-1991	0.366	0.333
	After 1991	0.259	0.311
Fraction of houses with area	Below 70 square meters	0.057	0.124
	70-110 square meters	0.315	0.215
	110-150 square meters	0.340	0.187
	Above 150 square meters	0.288	0.240
Fraction of houses	Owned by a housing association	0.184	0.192
Fraction of households	Single people household	0.341	0.142
	Two people household	0.309	0.075
	Household with children	0.349	0.114
Fraction of people of age	0-45 years	0.504	0.114
	45-65 years	0.287	0.064
	Above 65 years	0.207	0.098
	Logarithm of income	10.165	0.171
Urbanization level	Very strongly urbanized	0.172	0.377
	Strongly urbanized	0.218	0.413
	Moderately urbanized	0.161	0.367
	Slightly urbanized	0.173	0.379
	Not urbanized	0.275	0.447
Distance to	Very strongly urbanized neighborhood	26.777	24.949
Province	Limburg	0.067	0.249
	Drenthe	0.032	0.177
	Noord-Holland	0.161	0.368
	Gelderland	0.129	0.335
	Friesland	0.053	0.225
	Zuid-Holland	0.181	0.385
	Overijssel	0.072	0.259
	Groningen	0.036	0.188
	Noord-Brabant	0.148	0.257
	Utrecht	0.071	0.257
	Zeeland	0.029	0.168
	Flevoland	0.019	0.137

In Table 2, the sample statistics of the explanatory variables are displayed. We present three main findings from Table 2. First, we see that on average, a neighborhood consists of 36.2

percent of rowhouses, which makes rowhouses making up the largest share, while the smallest share is for apartments. On average, neighborhoods consist for only 10 percent of apartments.

Second, on average the largest fraction of houses in Dutch neighborhoods consist of houses built between 1965 and 1991, and of houses with an area between 110 and 150 square meters. This tells us that the average Dutch neighborhood for the largest part consists of mid-size houses and houses built in the late 1900's.

Furthermore, we see that only 17.2 percent of the neighborhoods has urbanization level 1, while almost half (47.7 percent) of the neighborhoods is in a city. This shows that not all neighborhoods that are in a city, have the highest urbanization level.

In Figures 16 to 20, the mean of the dependent variables in each urbanization category is plotted. Here, 1 stands for 'very strongly urbanized' and 5 stands for 'not urbanized'. Especially for the probability of having mental health problems, there is a clear difference between the highest urbanization level and the lowest urbanization level. This pattern not the same for the physical health problems. Here we see a uniform distribution across the urbanization levels, except for the lowest urbanization level, not urbanized. This motivates the question whether urbanization has a relationship with the probability of having health problems.

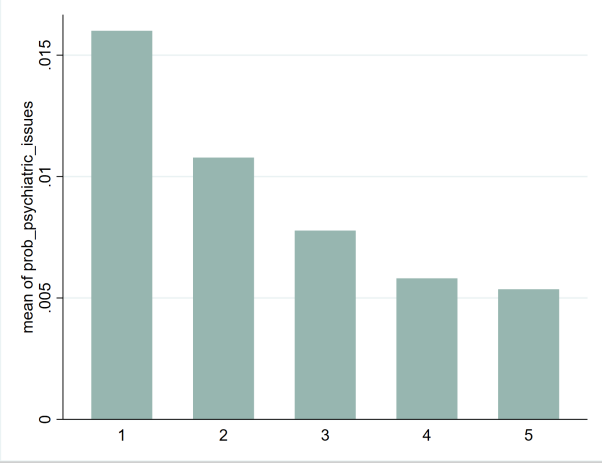

Figure 16: Mean probability of having psychiatric problems per urbanization level

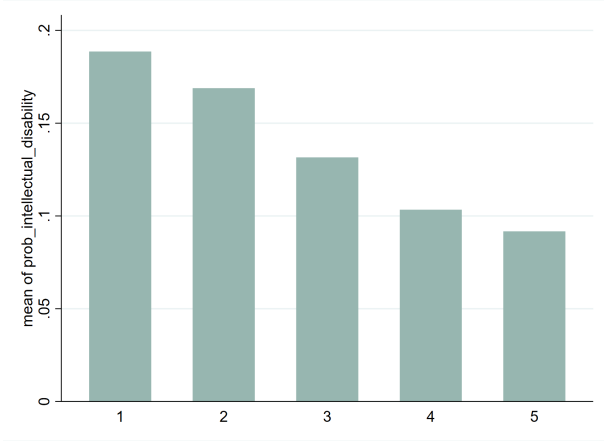

Figure 17: Mean probability of having an intellectual disability per urbanization level

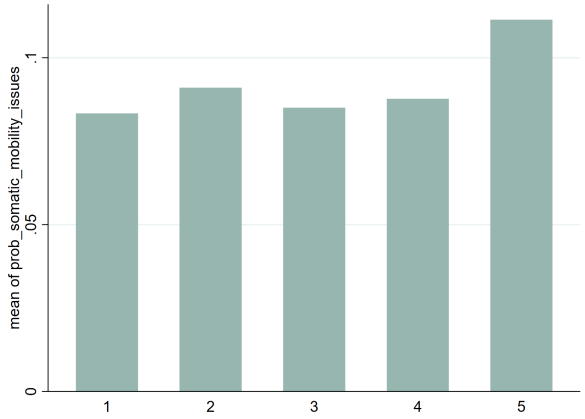

Figure 18: Mean probability of having somatic problems per urbanization level

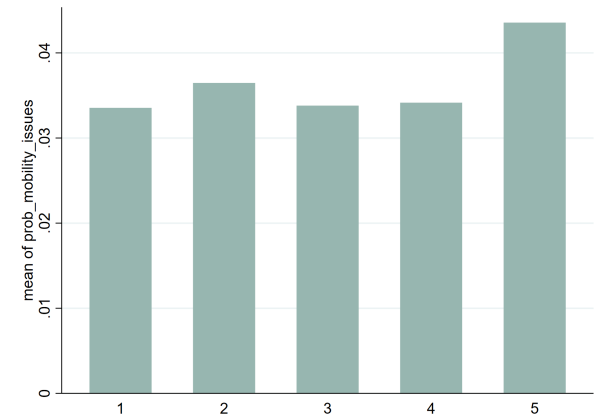

Figure 19: Mean probability of having mobility problems per urbanization level

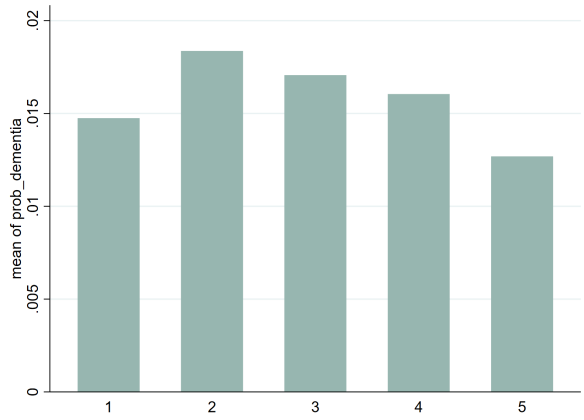

Figure 20: Mean probability of having dementia per urbanization level

Empirical approach

The aim of this paper is to explain the variation in health outcomes. We do this using the linear regression model, estimated by ordinary least squares (OLS). This estimation method minimizes the sum of squares of the differences between the values of the dependent variable and the values that are predicted by the linear function of the product of the independent variables and their parameters. In this way, possible relationships between independent variables and dependent variables can be investigated, in addition to the significance of these relationships.

The following equation represents the model:

$$y = X^\top \beta + \epsilon$$

y is the outcome variable. X is the matrix of explanatory variables and β is the vector of coefficients. ϵ is the error term. In the matrix of explanatory variables X, $x_1 = 1$ since β_0 is the constant term. We impose the following assumption on the error term ϵ:

$$E(\epsilon|X) = 0$$

We also assume that we have a random sample.

Since we have five different outcome variables, we use this model five times, one time for each of the following outcomes: the probability of having psychiatric problems, the probability of having an intellectual disability, the probability of having somatic problems, the probability of having mobility problems and the probability of having dementia.

For each outcome variable, the matrix X consists of the same explanatory variables. This leads to five similar regressions, where the only difference is the outcome variable. We do this to establish relationships between the outcome variables and all explanatory variables.

To establish a causal relationship between the urbanization level of a neighborhood and the outcome variables, we considered an instrumental variable (IV) regression. As instruments we constructed the travel distance from each neighborhood to a very strongly urbanized neighborhood and the building density in a neighborhood. We expected this to be correlated with the dummy for being a very strongly urbanized neighborhood, and this was the case. This makes both instruments relevant. However, the Sargan-test of over-identifying restrictions showed that

the instruments were not exogenous. Therefore, we can not take an IV approach. For more details on the IV procedure, see Appendix B.

4 Results

In this section we present results from linear regression models for the five health outcome variables.

Mental health problems

Table 3: Linear regression explaining the probability of having mental health problems

		Probability of having psychiatric problems	Probability of having an intellectual disability
Fraction of houses of type	Detached	0.000	−0.012
		(0.000)	(0.010)
	Semi-detached	−0.002**	−0.061***
		(0.001)	(0.010)
	Corner	−0.001	−0.024
		(0.001)	(0.015)
	Row	0.002***	−0.007
		(0.001)	(0.008)
Fraction of houses of construction year	Before 1945	0.003***	0.023***
		(0.000)	(0.004)
	1945-1965	0.003***	0.049***
		(0.000)	(0.005)
	1965-1991	0.001***	0.025***
		(0.000)	(0.003)
Fraction of houses of area	Below 70 sq. meters	−0.002***	0.014
		(0.001)	(0.014)
	70-110 sq. meters	0.000	0.065***
		(0.000)	(0.007)
	110-150 sq. meters	−0.001***	0.037***
		(0.000)	(0.008)
Fraction of houses	owned by housing association	0.011***	0.185***
		(0.001)	(0.007)
Fraction of households	Two people househ.	−0.036***	−0.269***
		(0.001)	(0.017)
	Househ. with children	−0.028***	−0.028*
		(0.001)	(0.014)
Fraction of people of age	45-65 years	0.006***	0.076***
		(0.001)	(0.017)
	Above 65 years	−0.004***	0.007
		(0.001)	(0.015)
Logarithm of income		−0.008***	−0.180***
		(0.000)	(0.007)
City		0.000**	−0.003

		(0.000)	(0.002)
Urb. level	Very strongly urbanized	0.002***	0.029***
		(0.000)	(0.004)
	Strongly urbanized	0.001***	0.026***
		(0.000)	(0.003)
	Moderately urbanized	0.001***	0.015***
		(0.000)	(0.002)
	Slightly urbanized	0.000***	0.007***
		(0.000)	(0.002)
Distance to	Very strongly urb. neighborhood	0.000**	0.000**
		(0.000)	(0.000)
Province	Limburg	−0.001***	0.022***
		(0.000)	(0.005)
	Drenthe	−0.001*	0.002
		(0.000)	(0.006)
	Noord-Holland	−0.005***	−0.013***
		(0.000)	(0.005)
	Gelderland	−0.003***	−0.012**
		(0.000)	(0.005)
	Friesland	−0.002***	−0.007
		(0.000)	(0.005)
	Zuid-Holland	−0.005***	−0.007
		(0.000)	(0.005)
	Overijssel	−0.002***	−0.005
		(0.000)	(0.005)
	Noord-Brabant	−0.003***	−0.001
		(0.000)	(0.005)
	Utrecht	−0.005***	−0.017***
		(0.000)	(0.005)
	Zeeland	−0.003***	−0.003
		(0.000)	(0.006)
	Flevoland	−0.001***	0.021***
		(0.000)	(0.006)
R-squared		0.723	0.622

***, ** and * denote statistical significance at the 1, 5, and 10 percent level, respectively. Standard errors between parentheses.

In Table 3 the results from a linear regression model are presented. The model is estimated by the OLS method. The outcomes are the probability of having psychiatric problems and the probability of having an intellectual disability. We present three main findings from these regressions.

The explanatory variables on house type, construction year and area of a house proxy for house quality. The probability of having mental health problems is in general lower when the fraction of detached, semi-detached, corner or rowhouses is larger in a neighborhood. This indicates that people living in an apartment have more mental health problems than other people. Also, the older the house is, the higher the probability of having mental health issues. The probability of having an intellectual disability is larger when the living area of a house is smaller, as the base case in the regression is the fraction of houses with an area above 150 square meters and the coefficients of the area variables are positive. If we assume that the older and smaller the house is, the worse its quality is, then the probability of having mental health

problems is higher when the quality of a house is bad, even when we control for income. These results are in line with what we find in literature. For instance, Evans et al. (2003) have shown that house quality has a is positively associated with psychological well-being.

The more houses are owned by a housing association in a neighborhood, the higher the probability of having mental health issues, even when controlling for income. This indicates that in municipalities where the social dwellings are highly concentrated in some neighborhoods, in these neighborhoods there may be a relatively large amount of people with mental health problems, since the fraction of social dwellings there is high.

The city indicator and the variables indicating urbanization level show a clear pattern. First, although significant at the 5 percent level, the city dummy is close to zero in the regression of psychiatric problems. It is insignificant in the regression of the probability of having an intellectual disability. This indicates that when we are already controlling for the urbanization level, the city dummy does not explain much variation in the outcome variables anymore.

The coefficients of the urbanization level dummies are positive and significant. The base variable here is 'not urbanized', the lowest urbanization level. The positive coefficients show that the more urbanized a neighborhood is, the higher the probability of having mental health problems is. This is also what Srivastava (2009) shows: urbanization is associated with an increase in mental disorders.

The distance to a very strongly urbanized neighborhood is significant, and the effect is close to zero but positive. This indicates that the people who live closely to a very strongly urbanized neighborhood have a higher probability of having mental health problems than the people who do not.

Especially in the regression of psychiatric problems, all province indicators are significant. The coefficients of the significant indicators are all negative, except for the Flevoland dummy and the Limburg dummy in the regression of the probability of having an intellectual disability. The base variable here is the Groningen dummy. This indicates that in general, people who live in any other region than Groningen have a smaller probability of having mental health problems. A potential explanation is that Groningen is a remote area. People go to other parts of the country to study or explore career opportunities where private and governmental organizations are located. It may be the case that people with mental health problems do not do this, so they prefer to reside in Groningen.

Physical health problems

In this section, we discuss the results of the OLS regressions of both the probability of having somatic problems and the probability of having mobility problems. The results are displayed in Table 4.

Table 4: Linear regression explaining the probability of having physical health problems

		Probability of having somatic problems	Probability of having mobility problems
Fraction of houses of type	Detached	0.06***	0.022***
		(0.007)	(0.002)
	Semi-detached	0.014*	0.006***
		(0.008)	(0.003)
	Corner	−0.031***	−0.007**
		(0.009)	(0.003)

	Row	0.015***	0.006***
		(0.005)	(0.002)
Fraction of houses of construction year	Before 1945	0.011***	0.004***
		(0.003)	(0.001)
	1945-1965	0.019***	0.007***
		(0.003)	(0.001)
	1965-1991	−0.001	−0.001
		(0.002)	(0.001)
Fraction of houses of area	Below 70 sq. meters	0.005	0.000
		(0.008)	(0.002)
	70-110 sq. meters	0.021***	0.008***
		(0.005)	(0.002)
	110-150 sq. meters	0.019***	0.006***
		(0.005)	(0.002)
Fraction of houses	Owned by housing association	0.061***	0.025***
		(0.003)	(0.001)
Fraction of households	Two people househ.	−0.149***	−0.062***
		(0.012)	(0.005)
	Househ. with children	−0.038**	−0.021**
		(0.009)	(0.003)
Fraction of people of age	Between 45-65 years	0.071***	0.025***
		(0.013)	(0.005)
	Above 65 years	0.214***	0.083***
		(0.010)	(0.004)
Logarithm of income		−0.008***	−0.018***
		(0.000)	(0.001)
City		−0.002**	−0.001**
		(0.001)	(0.000)
Urb. level	Very strongly urbanized	−0.017***	−0.007***
		(0.003)	(0.001)
	Strongly urbanized	−0.015***	−0.006***
		(0.002)	(0.001)
	Moderately urbanized	−0.015***	−0.006***
		(0.002)	(0.001)
	Slightly urbanized	−0.013***	−0.006***
		(0.001)	(0.001)
Distance to	Very strongly urb. neighborhood	0.000***	0.000***
		(0.000)	(0.000)
Province	Limburg	−0.003	−0.002
		(0.003)	(0.001)
	Drenthe	0.002	0.000
		(0.004)	(0.002)
	Noord-Holland	−0.007**	−0.005***
		(0.003)	(0.001)
	Gelderland	−0.008**	−0.004***
		(0.003)	(0.001)
	Friesland	−0.012***	−0.006***
		(0.003)	(0.001)
	Zuid-Holland	−0.007**	−0.006***

	(0.003)	(0.001)
Overijssel	−0.007**	−0.003***
	(0.003)	(0.001)
Noord-Brabant	−0.012***	−0.006***
	(0.003)	(0.001)
Utrecht	−0.010***	−0.006***
	(0.003)	(0.001)
Zeeland	−0.010**	−0.007***
	(0.004)	(0.001)
Flevoland	−0.007*	−0.004***
	(0.004)	(0.002)
R-squared	0.338	0.379

***, ** and * denote statistical significance at the 1, 5, and 10 percent level, respectively. Standard errors between parentheses.

As explained in the Canon Volkshuisvesting (2015), housing associations were established in the early 1900's to improve the health of the working class. However, the past years the focus of the housing associations has mainly been on making houses sustainable, building and renovating houses and keeping the houses affordable. The health of the people living in social dwellings is not the main focus anymore. To help housing associations shift the focus back to the health of the people living in the social dwellings, we analyze physical health problems in this paper.

In general, the probability of having physical health problems is higher in a neighborhood where the fraction of detached houses, semidetached houses and rowhouses is larger, in comparison to the fraction of apartments in a neighborhood. We would expect living in an apartment to have a negative impact on physical health, but when we control for area and income among other factors, this does not seem to be the case. All significant construction year and area variables have positive coefficients.

The construction year variables and area variables show a clear pattern. The older the house, the higher the probability of having physical health problems, as the base variable here is the fraction of houses with a construction year after 1991. The smaller the house, the higher the probability of having physical health problems as well, as in this case the base variable is the fraction of houses with an area larger than 150 square meters. These variables can again be a proxy for house quality, indicating that the probability of having physical health problems is larger in small and old houses, which we define as bad quality houses. This is what is shown in literature as well. Every additional poor quality housing characteristic leads to poorer health status (Boch et al., 2020).

The effect of the fraction of houses owned by a housing association is positive and significant, just as in the regressions of the probability of having mental health problems. This indicates that in neighborhoods where there are relatively many social dwellings, people have more physical health problems, even when we keep all variables, including income, constant.

Living in a city and living in urbanized areas have a negative and significant effect on the probability of having physical health problems, compared to not living in a city and living in a not urbanized neighborhood. This effect is the opposite of the effect we saw in the regressions of the probability of having mental health problems. It indicates that in highly urbanized areas, people in general have less physical health problems. We cannot establish causal effects in this paper, so it may be the case that people with physical health problems need more space to live in and therefore move out of an urbanized area, into a less urbanized area. The significant province

dummies all have negative effects, which means that living in any other area than Groningen is negatively related to the probability of having physical health problems, when keeping other variables constant. This may indicate that people in Groningen have more physical health problems than anywhere else in the Netherlands.

Dementia

In this section, we discuss results for the probability of having dementia. The results are presented in Table 5.

Table 5: Linear regression explaining the probability of having dementia

		The probability of having dementia
Fraction of houses of type	Detached	−0.004***
		(0.001)
	Semi-detached	0.000
		(0.001)
	Corner	0.002
		(0.002)
	Row	−0.002
		(0.001)
Fraction of houses of construction year	Before 1945	0.000
		(0.001)
	1945-1965	0.006***
		(0.001)
	1965-1991	0.002***
		(0.000)
Fraction of houses of area	Below 70 sq. meters	−0.005***
		(0.001)
	70-110 sq. meters	−0.001
		(0.001)
	110-150 sq. meters	−0.002**
		(0.001)
Fraction of houses	Owned by a housing association	0.000
		(0.001)
Fraction of households	Two people househ.	−0.017***
		(0.003)
	Househ. with children	−0.002
		(0.002)
Fraction of people of age	45-65 years	−0.008*
		(0.004)
	Above 65 years	0.080***
		(0.004)
Logarithm of income		0.002***
		(0.001)
City		−0.001***

		(0.000)
Urb. level	Very strongly urbanized	0.003***
		(0.001)
	Strongly urbanized	0.003***
		(0.000)
	Moderately urbanized	0.002***
		(0.000)
	Slightly urbanized	0.002***
		(0.000)
Distance to	Very strongly urb. neighborhood	0.000
		(0.000)
Province	Limburg	0.001
		(0.001)
	Drenthe	0.001
		(0.001)
	Noord-Holland	0.001
		(0.001)
	Gelderland	0.001*
		(0.001)
	Friesland	0.001*
		(0.001)
	Zuid-Holland	0.001
		(0.001)
	Overijssel	0.001**
		(0.001)
	Noord-Brabant	0.001
		(0.001)
	Utrecht	0.001
		(0.001)
	Zeeland	0.003***
		(0.001)
	Flevoland	0.000
		(0.001)
R-squared		0.572

***, ** and * denote statistical significance at the 1, 5, and 10 percent level, respectively. Standard errors between parentheses.

In 1990, only 12.8 percent of the total population in the Netherlands was above 65 years old. In 2020 already 19.5 percent of the total population is older than 65 years (CBS, 2020). The Dutch population is increasingly aging. This pressures the public expenditures, as the aging population implies that health care expenditures are increasing as well (ESB, 2018). Expenditures in healthcare for elderly people will grow from 17 billion in 2015 to 43 billion in 2040 (RIVM, 2018). 1 in 3 women and 1 in 7 men get dementia, mostly after the age of 65 years (Alzheimer Nederland, 2021). We analyze dementia in this paper to explore the relationship of the probability of having dementia and the explanatory variables.

The fraction of detached houses in the neighborhood is the only significant variable among the house type variables. Its coefficient is negative, indicating that the probability of having dementia is lower in neighborhoods with a large fraction of detached houses. The higher the

fraction of old houses, with a construction year between 1945 and 1991, the higher the probability of having dementia. Furthermore, the smaller the house is in terms of area, the smaller the probability of having dementia. This indicates that people living in a small and relatively new house are less likely to suffer from dementia. An explanation for this could be that people with dementia are elderly people who have moved into a large house with enough room for their possible children when they were younger, and still live there.

In the regressions of the probability of having mental and physical health problems, the fraction of houses owned by a housing association was significant. However, in the regression of the probability of having dementia, this is not the case. The variable is insignificant and its effect is close to zero. Furthermore, the income variable was significant and had a negative effect in the regressions of the probability of having mental and physical health problems. Here, we see the opposite: the effect of income is positive. This indicates that the people who suffer from dementia have a higher income. This can be an explanation for the insignificance of the fraction of houses owned by a housing association. People who suffer from dementia and live at home (since the health outcomes only consider people who do not live in a care facility but at home) are wealthier people, who probably do not qualify to live in a social dwelling, which makes the fraction of social dwellings in a neighborhood insignificant.

The variables that are indicators for the urbanization level of a neighborhood show that people with dementia are more likely to live in an urbanized area, outside of a city. This is contradictory to what we find in literature: Robbins et al. (2019) have found that rates of dementia among people living in urban areas are lower than people living in rural areas. However, in that research they included people living in a care facility, whereas in this research we only look at people living independently. The regional variation we detected in the regressions of the probability of having mental and physical health problems is not visible in the regression of the probability of having dementia. Except for a small number of province indicators, all province indicators are insignificant. This tells us that there is less regional variation in the probability of having dementia, when compared to the probability of having other mental and physical health issues.

In Table 7, 8 and 9 in Appendix A we show the results of the regressions of all outcome variables discussed before, but we omitted the province indicators. The effects of all other explanatory variables do not differ drastically from the effects that we have seen in Table 3, 4 and 5. This indicates that regional variation is significant, but that the effects of the other explanatory variables within the Dutch provinces are similar to the effects across the entire Netherlands.

Generally, the construction year variables are significant in the regressions for the probability of having mental and physical health issues and dementia. This has raised the question whether there is a significant difference in the effect of the construction year in a very strongly urbanized neighborhood and outside a very strongly urbanized neighborhood. In Table 6 the results of five regressions, one of each outcome variable discussed before, are displayed. In this regression we included the baseline explanatory variables, and interaction terms between the construction year variables and the indicator for the highest urbanization level. The interaction terms are significant in the regressions of the probability of having an intellectual disability, somatic problems and mobility problems. They show that in a very strongly urbanized neighborhood the effect of the fraction of the houses that have a construction year before 1991 is less positive than outside a very strongly urbanized neighborhood, or even negative. We can assume that the construction year of a house is a proxy for the quality of a house. Then, this indicates that in a very strongly urbanized neighborhood, it is not necessarily the case that the more bad quality houses in a neighborhood there are, the worse the mental or physical health in the neighborhood is.

Table 6: Results of OLS regression

	Probability of having psychiatric problems	Probability of having an intellectual disability	Probability of having somatic problems	Probability of having mobility problems	Probability of having dementia
Constr. year below 1945	0.002***	0.027***	0.025***	0.008***	0.000
	(0.000)	(0.005)	(0.004)	(0.001)	(0.000)
Constr. year 1945-1965	0.003***	0.052***	0.018***	0.007***	0.006***
	(0.000)	(0.006)	(0.004)	(0.001)	(0.001)
Constr. year 1965-1991	0.001***	0.028***	0.000	0.000	0.002***
	(0.000)	(0.003)	(0.002)	(0.001)	(0.000)
Constr. year below 1945	0.002***	-0.023**	-0.040***	-0.013***	0.001
× very strongly urbanized	(0.001)	(0.010)	(0.005)	(0.002)	(0.001)
Constr. year 1945-1965	0.000	-0.025**	-0.010**	-0.003***	0.000
× very strongly urbanized	(0.001)	(0.011)	(0.005)	(0.002)	(0.001)
Constr. year 1965-1991	0.001	-0.026***	0.000	0.001	0.002**
× very strongly urbanized	(0.000)	(0.002)	(0.004)	(0.002)	(0.001)

Baseline variables are not included in this table, but are controlled for in the regression. The construction year variables all concern the fraction of houses in the neighborhood with a construction year in that particular category. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level, respectively. Robust standard errors between parentheses.

5 Conclusion and discussion

The aim of this paper is to explain health outcomes across the Netherlands using data on the (built) environment, household characteristics and regional indicators. We divided the health outcomes into three categories: mental health problems, physical health problems and dementia.

In all three categories we aimed at explaining health outcomes using the same explanatory variables. These variables consisted of information on houses, such as house type, construction year and area, and information on households, such as household composition, age and income. We also included explanatory variables to account for regional variation, such as province dummies and indicators for the urbanization level of neighborhoods.

The results show that the probability of having physical and mental health problems is larger in neighborhoods where the fraction of small and old houses, which we define as the fraction of bad quality houses, is large. This result indicates that housing associations may benefit from building new houses or renovating bad-quality houses. However, the effect of the construction year of the houses is less strong in very strongly urbanized neighborhoods than outside such neighborhoods. This indicates that within very strongly urbanized neighborhoods, a larger fraction of bad quality houses does not necessarily indicate a large probability of having mental or physical health problems. This means that housing associations should start by focusing on renovating and building new houses in less urbanized areas to improve health in an efficient manner. Since the results also show that the larger the fraction of social dwellings in a neighborhood is, the more mental and physical health problems there may be, it might be good to create diversity in the neighborhood. This means that municipalities and housing associations should make sure that each neighborhood consists of a mix of social dwellings and other houses. In this way, the fraction of houses of a housing association is lower, which may decrease the probability of having mental and physical health problems in the neighborhood.

We found regional variation in all outcome variables. For the probability of having mental health problems and the probability of having dementia we found that the more urbanized a neighborhood is, the higher the probability of having mental health problems is. This implies that for municipalities to improve the mental health status of their inhabitants, it is good to monitor people that live in (very) strongly urbanized neighborhoods. Organizing activities or deploying so-called 'neighborhood-coaches' to be a point of contact for people may improve the mental health status of people. This also might prevent people suffering from dementia from being lonely.

For the probability of having physical health issues we found the opposite pattern. In strongly urbanized neighborhoods the probability of having physical health issues is lower than in less urbanized neighborhoods. This relationship could work two ways. Either people that live in less urbanized areas have less access to health care, which leads to more physical health issues. On the other hand, people who have physical health issues need enough space to live and may move to less urbanized areas for that reason. This implies that it is important for municipalities and housing associations to renovate or build good-quality houses in less urbanized areas, so people with physical health issues can live comfortably. Doing this may in the long term lead to less public expenditures, as people with physical health issues will be able to live at home instead of a care facility for a longer time, which can decrease the healthcare expenditures.

We find variation across the Dutch provinces for the probability of having mental health problems and the probability of physical health problems. For both probabilities we found that, in general, people in any other province than Groningen have less physical and mental health problems. We can interpret this in the following way: Groningen is a remote area that is far away from big cities in Randstad, for example. Study and career opportunities are generally better

somewhere else in the Netherlands than Groningen. People without any mental or physical health problems will move out of Groningen for their study or work, since they have more opportunities elsewhere. This leaves relatively many people with mental and physical health issues in Groningen, which could explain the pattern we see. However, we cannot establish a causal relationship, so this is merely speculation.

The analyses in this paper are restricted in several respects. First, we are not able to establish causal relationships. We aimed at performing an instrumental variable analysis (see Appendix B), but we were not able to find suitable instruments. Therefore, we cannot make a causal claim in this paper.

Second, in all models we see that the explanatory variables do not fully explain the variation in the outcome variables. We especially see this in the models for physical health: the R-squared in those models never exceeds 40 percent. This means that there are a lot of other factors, probably among which genetic factors, that explain the variation in the health outcomes. When we would have more data on these factors, we would be able to improve our models and explain more variation than we do at the moment.

Next, in this paper we aggregated the data to a neighborhood level. In the data almost 10,000 neighborhoods are represented. Some neighborhoods were not included in the analysis because of missing data. However, the models would be better if all neighborhoods were represented in the sample data models. What would improve the models even more is if we would have had data available on a finer level. At the moment, almost all independent variables were only available on a neighborhood level, but it would lead to better results if we would for instance have the data available on household-level. This would mean that for each house we would like to know what household lives in the house, what age they have and what income they earn. However, because of privacy matters, this will not be possible.

In this paper we have data available for the entire Netherlands. Therefore, we can state that the results we obtain can be interpreted for the Netherlands. However, it might be complicated for someone outside of the Netherlands to interpret these results. Every country has its own culture, its own housing infrastructure and its own healthcare system. A limitation of this paper might be that the implications of the results may not be extended to other countries.

A recommendation for future work is to try to establish causal effects. In this thesis, we attempted to use an instrumental variable regression, but we did not succeed in finding the right instruments. However, when there are both more resources and time available, it could be possible to find the right instruments. In this way, we can really establish causal effects. In future research it could also be possible to collect data on a more detailed level than the neighborhood level. In this way, the effects we found might be more robust.

Another recommendation would be to extend the research to more countries than only the Netherlands. It would be interesting to see whether the effects we find in this thesis are also applicable in other countries or if the effects are very country-specific.

Acknowledgements

I would like to thank everyone who made this paper a possibility and a success. My gratitude goes out to Dr. Tunga Kantarci, for his interest in and enthusiasm about this paper and his academic support. This paper would have not been a success without his advice and his feedback. It has been a joy to work with such an involved supervisor. I would also like to thank Republiq for providing me with an interesting research question and the rich data set I got to work with. I especially want to thank Peter Korst for motivating me when I needed it, for the good conversations we have had and will continue to have, and for trusting me. I would also like to thank my other colleagues from Republiq, especially Sara Rieuwerts, Leoniek Driessens and Alex Veraart, for inspiring and helping me. Lastly I want to express my gratitude to my friends and family, for the mental support they gave me throughout the past years of being a student.

References

Adeline, A., Delattre, E. (2017). Some microeconometric evidence on the relationship between health and income. *Health Economics Review, volume 7, issue 27.*

Adler, N., Ostrove, J. (1999). Socioeconomic Status and Health: What We Know and What We Don't. *Annals of the New York Academy of Sciences, volume 896, issue 1.* p. 3-15.

Alzheimer Nederland (2021). Factsheet cijfers en feiten over dementie.

Boch, S.J., Taylor, D.M., Danielson, M.L., Chisolm, D.J., Kelleher, K.J. (2020). 'Home is where the health is': Housing quality and adult health outcomes in the Survey of Income and Program Participation. *Preventive Medicine, volume 132.*

Cacioppo, J.T., Hawkley, L.C., Ernst, J.M., Burleson, M., Berntson, G.G., Nouriani, B., Spiegel, D. (2006). Loneliness within a nomological net: An evolutionary perspective. *Journal of Research in Personality, volume 40, issue 6.* p.1054-1085.

Canon Volkshuisvesting (2016). Retrieved from: https://www.canonsociaalwerk.eu/0000_boekvolkshuisvesting/Canon_volkshuisvesting_site.pdf

CBS (2020). Hoeveel ouderen zijn er in Nederland? Retrieved from: https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/leeftijd/ouderen

ESB (2020). Vergrijzing zet in gehele EU druk op overheidsfinanciën. Retrieved from: https://esb.nu/kort/20042396/vergrijzing-zet-in-gehele-eu-druk-op-overheidsfinancien

Evans, G.W., Wells, N.M., Moch, A. (2003). Housing and Mental Health: A Review of the Evidence and a Methodological and Conceptual Critique. *Journal of Social Issues, volume 59, issue 3.* p. 475-500.

Fuller, T.D., Edwards, J.N., Sermsri, S., Vorakitphokatorn, S. (1993). Housing, stress and physical well-being: Evidence from Thailand. *Social Science Medicine, volume 36, issue 11.* p. 1417-1428.

Habib, R.R., Mahfoud, Z., Fawaz, M., Basma, S.H., Yeretzian, J.S. (2009). Housing quality and ill health in a disadvantaged urban community. *Public Health, volume 123, issue 2.* p. 174-181.

Hawkley, L.C., Cacioppo, J.T. (2010). Loneliness Matters: A Theoretical and Empirical Review of Consequences and Mechanisms. *Annals of Behavioral Medicine, volume 40, issue 2.* p. 218-227.

Johnson, W., Krueger, R.F. (2005). Genetic Effects on Physical Health: Lower at Higher Income Levels. *Behavior Genetics, volume 35.* p. 579-590.

Jorm, A.F. (2000). Does old age reduce the risk of anxiety and depression? A review of epidemiological studies across the adult life span. *Cambridge University Press Psychological*

Medicine, volume 30, issue 2. p. 11-22.

Leidelmeijer, K., van Rossum, F., van Iersel, J. (2016). Ontwikkeling Woonzorgwijzer, Kwantificering groepen met beperkingen.

Lorem, G.F., Schirmer, H., Wang, C.E.A., Emaus, N. (2017). Ageing and mental health: changes in self-reported health due to physical illness and mental health status with consecutive cross-sectional analyses. *BMJ Open, volume 7, issue 1.*

McCarthy, D., Saegert, S. (1978). Residential Density, Social Overload and Social Withdrawal. *Human Ecology, volume 6, issue 3.* p. 253-272.

RIVM (2018). Zorgkosten stijgen naar 174 miljard in 2040. Retrieved from: https://www.zorgvisie.nl/rivm-zorgkosten-stijgen-naar-174-miljard-in-2040/

Robbins R.N, Scott, T., Joska, J.A., Gouse, H. (2019). Impact of urbanization on cognitive disorders. *Current Opinion in Psychiatry, volume 32, issue 3.* p. 210-217.

Shaw, M. (2004). Housing and public health. *Annual Review of Public Health, volume 25.* p. 397-418.

Smith, K.J., Victor, C. (2019). Typologies of loneliness, living alone and social isolation, and their associations with physical and mental health. *Cambridge University Press Ageing and Society, volume 29, issue 8.* p. 1709-1730.

Srivastava, K. (2009). Urbanization and mental health. *Industrial Psychiatry Journal, volume 18, issue 2.* p. 75-76.

Stordal, E., Mykletun, A., Dahl, A.A. (2003). The association between age and depression in the general population: a multivariate examination. *Atca Psychiatrica Scandinavica, volume 107, issue 2.* p. 132-141.

Sundquist, K., Ahlen, H. (2006). Neighbourhood income and mental health: A multilevel follow-up study of psychiatric hospital admissions among 4.5 million women and men. *Health and Place, volume 12, issue 4.* p. 594-602.

A Models without province dummy's

Table 7: Linear regression explaining the probability of having mental health problems

		The probability of having psychiatric problems	The probability of having an intellectual disability
Fraction of houses of type	Detached	0.001	−0.019**
		(0.001)	(0.010)
	Semi-detached	0.000	−0.063***
		(0.001)	(0.010)
	Corner	−0.000	−0.030**
		(0.001)	(0.015)
	Row	0.002***	−0.006
		(0.001)	(0.008)
Fraction of houses of construction year	Before 1945	0.003***	0.023***
		(0.000)	(0.004)
	1945-1965	0.002***	0.053***
		(0.000)	(0.005)
	1965-1991	0.001***	0.026***
		(0.000)	(0.003)
Fraction of houses of area	Below 70 sq. meters	−0.005***	−0.004
		(0.001)	(0.014)
	70-110 sq. meters	−0.001***	0.049***
		(0.000)	(0.007)
	110-150 sq. meters	−0.002***	0.030***
		(0.000)	(0.007)
Fraction of houses	Owned by housing association	0.012***	0.187***
		(0.001)	(0.007)
Fraction of households	Two people househ.	−0.037***	−0.263***
		(0.001)	(0.017)
	Househ.with children	−0.032***	−0.044***
		(0.001)	(0.013)
Fraction of people of age	Between 45-65 years	0.007***	0.091***
		(0.001)	(0.018)
	Above 65 years	−0.006***	0.000
		(0.001)	(0.001)
Logarithm of income		−0.008***	−0.194***
		(0.000)	(0.006)
City		−0.000	−0.003*
		(0.000)	(0.002)
Urb. level	Very strongly urbanized	0.001***	0.025***
		(0.000)	(0.004)
	Strongly urbanized	0.001***	0.024***
		(0.000)	(0.003)
	Moderately urbanized	0.001***	0.014***
		(0.000)	(0.002)
	Slightly urbanized	0.000	0.006***

		(0.000)	(0.002)
Distance to	Very strongly urb. neighborhood	0.000**	0.000**
		(0.000)	(0.000)
R-squared		0.696	0.614

***, ** and * denote statistical significance at the 1, 5, and 10 percent level, respectively. Standard errors between parentheses.

Table 8: Linear regression explaining the probability of having physical health problems

		The probability of having somatic problems	The probability of having mobility problems
Fraction of houses	Detached	0.059***	0.021***
		(0.007)	(0.003)
	Semi-detached	0.014**	0.007***
		(0.007)	(0.003)
	Corner	−0.033***	−0.008**
		(0.009)	(0.003)
	Row	0.014***	0.005**
		(0.005)	(0.002)
Fraction of houses of construction year	Before 1945	0.012***	0.004***
		(0.003)	(0.001)
	1945-1965	0.019***	0.007***
		(0.003)	(0.001)
	1965-1991	−0.001	−0.001
		(0.002)	(0.001)
Fraction of houses of area	Below 70 sq. meters	0.005	−0.001
		(0.008)	(0.003)
	70-110 sq. meters	0.020***	0.007***
		(0.004)	(0.002)
	110-150 sq. meters	0.020***	0.006***
		(0.005)	(0.002)
Fraction of houses	Owned by housing association	0.061***	0.026***
		(0.003)	(0.001)
Fraction of households	Two people househ.	−0.149***	−0.063***
		(0.012)	(0.005)
	Househ. with children	−0.038***	−0.023***
		(0.008)	(0.003)
Fraction of people of age	Between 45-65 years	0.077***	0.027***
		(0.013)	(0.005)
	Above 65 years	0.216***	0.083***
		(0.009)	(0.004)
Logarithm of income		−0.046***	−0.020***
		(0.003)	(0.001)
City		−0.002	−0.001**
		(0.001)	(0.000)
Urb. level	Very strongly urbanized	−0.017***	−0.008***
		(0.003)	(0.001)

	Strongly urbanized	−0.015***	−0.006***
		(0.002)	(0.001)
	Moderately urbanized	−0.015***	−0.006***
		(0.002)	(0.001)
	Slightly urbanized	−0.013***	−0.006***
		(0.001)	(0.001)
Distance to	Very strongly urb. neighborhood	0.000***	0.000***
		(0.000)	(0.000)
R-squared		0.338	0.373

***, ** and * denote statistical significance at the 1, 5, and 10 percent level, respectively. Standard errors between parentheses.

Table 9: Linear regression explaining the probability of having dementia

		The probability of having dementia
Fraction of houses of type	Detached	−0.004**
		(0.001)
	Semidetached	−0.000
		(0.003)
	Corner	0.003
		(0.002)
	Row	−0.002
		(0.001)
Fraction of houses of construction year	Before 1945	−0.000
		(0.001)
	1945-1965	0.006***
		(0.001)
	1965-1991	0.002***
		(0.000)
Fraction of houses of area	Below 70 sq. meters	−0.005***
		(0.001)
	70-110 sq. meters	−0.001
		(0.001)
	110-150 sq. meters	−0.002**
		(0.001)
Fraction of houses	Owned by a housing association	0.000
		(0.001)
Fraction of households	Two people househ.	−0.016***
		(0.003)
	Househ. with children	−0.002
		(0.002)
Fraction of people of age	45-65 years	−0.008**
		(0.004)
	Above 65 years	0.080***
		(0.004)

Logarithm of income		0.002***
		(0.001)
City		−0.001***
		(0.000)
Urb. level	Very strongly urbanized	0.002***
		(0.001)
	Strongly urbanized	0.003***
		(0.000)
	Moderately urbanized	0.002***
		(0.000)
	Slightly urbanized	0.002***
		(0.000)
Distance to	Very strongly urb. neighborhood	0.000
		(0.000)
R-squared		0.571

***, ** and * denote statistical significance at the 1, 5, and 10 percent level, respectively. Standard errors between parentheses.

B IV results

In order to establish a causal relationship between the urbanization level of a neighborhood and health problems, we used an instrumental variable approach. We started with trying to instrument the dummy for 'very strongly urbanized' in the model of the probability of having psychiatric problems. The instruments we used for this were the building density in a neighborhood, which indicates the amount of buildings there are in a neighborhood, and the travel distance in kilometers to a neighborhood that is very strongly urbanized.

In an instrumental variable model, we need the instruments to be both relevant and exogenous. Relevance means that they should be correlated with the variable we instrument. We expected the building density to be correlated with the 'very strongly urbanized' dummy, since it is easy to imagine that areas with a high number of buildings are the more urbanized areas. Next to this, the travel distance to a very strongly urbanized neighborhood is correlated with the 'very strongly urbanized' dummy, since it is equal to 0 if the 'very strongly urbanized' dummy is equal to 1. To see whether the instruments meet the relevance criterion, we regress the probability of having psychiatric issues on the building density and the travel distance to a very strongly urbanized neighborhood. This leads to the results in Table 10.

Table 10: Linear regression of 'very strongly urbanized' dummy on instruments

	Coefficients
Distance to very strongly urbanized neighborhood	−0.005***
	(0.000)
Building density	0.000***
	(0.000)

***, ** and * denote statistical significance at the 1, 5, and 10 percent level, respectively. Standard errors between parentheses.

As shown in Table 10, we see that both instruments are significant in the regression, which

means that they are relevant. The next criterion they should meet is the exogeneity criterion. This means that both instruments should not be correlated with the dependent variable. We test this using the Sargan test of over-identifying restrictions. This test rejects the null hypothesis that the instruments are exogenous. This means that we cannot use the instruments, since they do not meet the exogeneity criterion.

We retrieved the same results for the other dependent variables. Every time the instruments are relevant, but not exogenous. We also tried more combinations of instruments. For instance, we used the fraction of apartments in the neighborhood as an instrument, and the fraction of houses with a construction year below 1945. We expected those variables to be relevant, since in highly urbanized areas we find more apartments and older houses. Also these instruments were relevant, but they were not exogenous.