Ageing Hearts, Limited Care

The Impact of Reduced Access to Primary Healthcare on Cardiovascular Medications Rates among Older Adults in The Netherlands

(image generated with Midjourney)

Bachelor Thesis in Human Geography and Planning
Christina Bollmann
S4424743
University of Groningen - Faculty of Spatial Sciences
Supervisors: Tobias Vogt and Lara Bister
26th of January 2023

Abstract

The declining number of local general practitioner practices coincides with a growing demand for primary healthcare in an ageing population. This trend calls for an investigation to inform future policies to address the unmet primary healthcare needs of vulnerable demographics. The present thesis aims to assess the impact of declining general practitioner practices on dispensed prescriptions of cardiovascular medication in older adults in The Netherlands. The Data came from multiple databases by the Dutch *Central Bureau voor Statistics* (CBS), providing municipality-level data on the proximity to general practitioners(GPs), prescription medication rates for different age groups, and various socio-economic indicators. These were combined to build a Pooled Ordinary Least Squares regression model, incorporating data from 334 Dutch municipalities spanning six years, 2014-2019. Key findings reveal that a 1% increase in the proximity of general practitioners leads to a 0.3% increase in cardiovascular medication prescriptions, with more pronounced effects in rural areas. These results indicate that primary healthcare access is linked to treatment capabilities in the older population and thus could have significant implications for the older population's cardiovascular health. The impact of spatial and temporal factors on this relationship highlights the need for further nuanced research into healthcare accessibility and its effects on elderly populations.

Table of Contents

List of Figures and Tables

1. Introduction

1.1 Background

As the Netherlands confronts an ageing population and a rise in age-related chronic illness rates, there is a growing need for effective healthcare services (RIVM, 2023). At the same time, the primary healthcare system facing these increasing demands is strained by a growing shortage of general practitioners (ABN Amro, 2023), which may be leading to diminished access to care for many elderly individuals.

Particularly interesting are the effects on preventive cardiovascular care due to this pressure on healthcare accessibility. Age is an independent non-modifiable risk factor for cardiovascular disease (CVD), as the physical changes of ageing lead to stiffening arteries and reduced heart muscle efficiency (AHA, 2022). This is further exacerbated by the compound effects of other risk factors which are often themselves more prevalent with higher age, such as smoking, obesity, diabetes, and frailty (Rodgers et al., 2019). Though, compared to age, some of these are modifiable risk factors for which intervention can be effective at reducing future disease burden (Koopman et al., 2016).

While these modifiable risk factors can be addressed to lead to better health outcomes, individuals with high risk factors must first be identified. In the Dutch health care system, this regular risk assessment is usually done by general practitioners. Once an individual has been identified as at-risk, treatment options include lifestyle changes and medication treatment (NHG, 2019). Common risk factors for CVD include hypertension and high cholesterol, which are treated by medications classified under the Anatomical Therapeutic Chemical (ATC) code group C, alongside other medications that are used for the treatment of conditions of the cardiovascular system (WHO, 2023).

According to a recent estimate, cardiovascular disease cost the European Union €282 billion annually, including the cost of health and social care, informal care, and loss in productivity due to disability and premature death (Luengo-Fernandez et al. 2023). As such, it is important to ensure that adequate care can be provided, and to see whether there are spatial inequalities in the access of this care. As, notably, those residing in the rural regions are at greater risk, as they not only experience a decline in services associated with population loss, but are also the municipalities with the oldest average populations (CBS, 2023).

1.2 Research Aim

This thesis seeks to investigate the interplay and relationship between changes in the spatial accessibility to primary care practices and dispensed cardiovascular medication rates among the older adult population. The aim is to comprehend the broader implications of healthcare service losses within municipalities.

Central Research Question:

- "To what extent do changes in the availability of local primary healthcare practices influence changes in cardiovascular medication prescriptions in municipalities in the Netherlands?"

Secondary Questions:

- “Does the demographic composition, specifically the proportion of elderly residents, in a municipality influence the relationship between GP accessibility and cardiovascular medication prescriptions?”
- “How does the association differ between rural and urban municipalities?”

1.3 Structure of Thesis

First, the theoretical framework aims to summarise the existing literature relating to primary healthcare disparities, prescription rates as an indicator for population health, and the spatial and socioeconomic factors that play a role in cardiovascular health outcomes on a population level. This forms the basis of the conceptual model, which is used as a framework for the research design. Then, the data sources, transformations and operationalisation, and the analytical strategy are laid out in the ‘Methodology’ section. This section further includes a descriptive statistics table of the final sample. The findings are then presented and discussed, under the header ‘Results and Analysis’, which include summarising the sample, and the results table for the pooled OLS regression models, which are then first interpreted. This is followed by a critical discussion of the findings and the limitations of this study. Lastly, the ‘Conclusion’ section summarises the main points, and offers ideas for possible further research.

2. Theoretical Framework

2.1 Access to Primary Healthcare and Population Health Outcomes

Academic interest in spatial disparities concerning healthcare access has a long history, yet the literature regarding the impacts on prescription medication rates, are sparse. One study that mentions the relationship between spatial access to healthcare facilities and prescription medications is the analysis by Rushworth et al. (2017) in the rural Scottish Highlands. The survey responses showed that higher age was correlated with an increased difficulty in accessing primary healthcare practices. It also further found that those elderly respondents that required more than five prescription medications struggled more with accessing these medications.

Within an urban context, another European study (Padeiro, 2018) only focussing on the proximity to pharmacies in Lisbon, Portugal found that only around 60% of elderly residents live within a 10 minute walk of a pharmacy, though it does not discuss impacts on prescription rates or health outcomes, it shows that difficulties with spatial access are not only a rural problem.

<u>Hypothesis 1</u>: Changes in nearby GP practices are correlated to changes in the rates of cardiovascular prescription medication.

Outside of the European context, various studies from the US and Canada have taken different approaches to shed light on the relationship between contextual factors, health outcomes, and spatial disparities. Though, they focus on healthcare utilisation and health outcomes, rather than prescription medications. Nemet and Bailey (2000) explored the connection between proximity to healthcare facilities and utilisation among elderly residents in rural Vermont. They found a statistically significant decrease in primary healthcare utilisation with increasing distance from a general practitioner, but emphasised that distance merely serves as a surrogate for a more complex experience among older adults.

This importance of individual factors is supported by a neighbourhood-level study in Seattle, Goh et al. (2018) which delved into the impact of nearby health facilities on sudden cardiac arrest incidence. Their findings did not support the notion of a protective effect from locally available medical services; in fact, they even reported higher odds of sudden cardiac arrest, explained by individual socio-economic circumstances outweighing proximity.

Furthering our understanding of healthcare disparities related to the more complex socio-economic factors that play a role in healthcare access, Gilliland et al. (2019) employed a geospatial approach to assessing primary healthcare accessibility among vulnerable populations, including the elderly, single parents, linguistic minorities, and low-income groups, within the city of London, Ontario, Canada. They revealed significant challenges in meeting the healthcare needs of these groups, further emphasising the importance of addressing spatial disparities.

<u>Hypothesis 2</u>: The spatial context has an effect on the association between GP accessibility and medication rate changes.

Though the research on this topic is limited, it generally indicates that reduced access to healthcare facilities can lead to lowered utilisation of healthcare services, which may, in turn, affect the rate at which prescription medications are obtained. However, it's important to note that socio-economic status, age, and the specific geographic context also significantly influence these outcomes.

2.2 Prescription Medication Rates as a Marker of Population Health

In regards to the usage of prescription medication rates as a proxy outcome variable in a cross-sectional study, multiple studies in the past few years have used mental health medication prescription rates in combination with contextual or spatial factors. For example, Aerts et al. (2022) in Belgium explored the relationship between residential green space and mental health medication sales. Their findings revealed that increased green cover was associated with decreased medication sales, highlighting the role of the environment in mental health-related medication. Similarly, McDougall et al. (2021) investigated the impact of 'blue space,' referring to aquatic environments, on mental health prescription medication among older adults in Scottish neighbourhoods. They found that the presence of both freshwater and coastal environments in proximity was associated with lower antidepressant medication prescription rates.

Beyond these ecological studies, the impact of spatial factors on prescription medication rates is sparse, and as such, this study aims to further the understanding of the spatial disparities in medication rates as a cross-sectional comparison tool.

Despite the extensive research into healthcare access disparities, a gap exists in understanding their influence on prescription medication rates, especially for cardiovascular prescriptions, which become increasingly important as populations age. This becomes particularly relevant when considering the varied age distributions among Dutch municipalities, with peripheral regions experiencing a rise in elderly populations (CBS, 2023).

Considering the demographic distribution within the Netherlands, recent research by Vulpen et al. (2023) adds further context by exploring regional development trajectories. Their findings reveal a decline in public services in rural periphery areas, coinciding with the highest share of the population aged 65 and above in these regions. Therefore, we should expect a difference between rural and urban regions and the impact of declining GP practices on the prescription rates of cardiovascular medication.

These trends underscore the need to comprehensively address spatial disparities in healthcare access and medication prescriptions, particularly for cardiovascular health. Especially for the growing share of older adults, the decreased accessibility to primary healthcare facilities could lead to irregularities or lapses in prescriptions. Taken together, we would expect the share of residents over the age of 65 to impact the relationship between cardiovascular prescription rates and GP accessibility.

Hypothesis 3: The demographic composition of Dutch municipalities play a role in the relationship between GP accessibility and cardiovascular medication prescription rates.

However it is important to note that other factors influence the rate of prescription medication. For example, national awareness campaigns concerning preventive health, which have promoted physical fitness and smoking cessation. Koopman et al. (2016) states this contributed to the improvement of national cardiovascular health, and through that possibly also lowered the need for cardiovascular medication prescriptions.

2.3 Conceptual Model

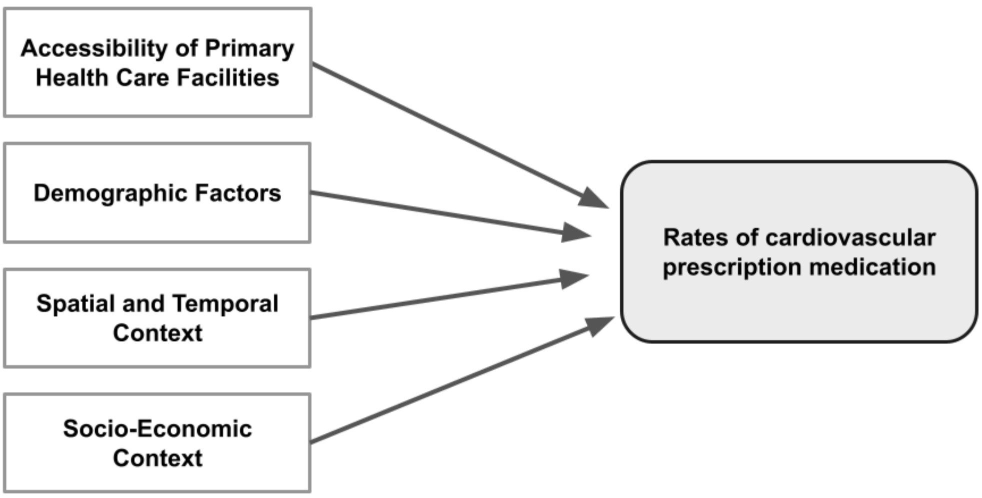


Figure 1: Conceptual Model

Building upon the theoretical framework, this conceptual model (Figure 1) seeks to illustrate the relationship between the variable of declining GP practices and the response variable of the rate of cardiovascular prescription medication. On the top is the factor of interest for this study, the accessibility of healthcare facilities, followed by the independent control factors.

Firstly, while the existing literature is limited, the influence of the age demographics of a municipality are important to control for, due to possible confounding factors of excess primary healthcare needs of an older population. Further, as studies found disparities when considering individual socio-economic circumstances (Gilliand et al., 2019; Nemet and Bailey's, 2000), it is important to also consider these effects in this study. Lastly, considering the broader context by controlling for the built environment and yearly changes is necessary to better account for not only the individual differences but also wider national trends and changes in prescription behaviour that could affect the outcome variable.

3. Methodology

3.1 Data Sources

For this analysis, five different datasets were utilised. Of this, four are CBS datasets (CBS SatLine, 2013-2019 a-d) and the fifth is data for the shapefiles of the municipal boundaries, which is maintained through cooperation of CBS and the Dutch government service *Public Services on the Map (*PDOK & CBS, 2014-2019*).* The following CBS datasets include longitudinal and cross-sectional data aggregated on the level of municipalities:

The datasets '83251NED: Personen met verstrekte geneesmiddelen' [*'Persons with prescription medication'*] and '80305ENG: Proximity to amenities', both from CBS StatLine (2013-2019a & 2013-2019b), provide the variables on the Dutch population's medication usage and the accessibility of essential services, respectively. While the former presents an annual overview of dispensed medication, excluding hospital and nursing home patients, categorised by the ATC Drug classification system, the latter gives figures on the spatial proximity of facilities by calculating road distances from residential addresses, providing a weighted average distance and facility count within specified radii.

Additionally, dataset 70072NED (CBS StatLine, 2013-2019c), 'Regional key figures,' presents a detailed compilation of regional statistics, including numbers on demographics, employment, housing and income. Complementing this, dataset 60039FVW (CBS StatLine, 2013-2019d), 'Measures of the Financial Relations Act,' informs governmental fund allocations, providing various fiscal indicators including categorisation of municipalities as predominantly rural or urban, managed by the Ministry of the Interior and Kingdom Relations.

Lastly, the Dutch government service *Public Services on the Map (PDOK)* provides the shapefiles dataset 'CBS Gebiedsindelingen' ['CBS *Area divisions*'] (PDOK & CBS, 2014-2019) for administrative boundaries at various spatial scales in cooperation with CBS.

3.1.1 Ethical Considerations

The study relies solely on secondary data, which has been cleaned, population-scaled, and stripped of any identifiable information. CBS's privacy policy ensures the data is free from privacy risks. Further information can be found in CBS's privacy statement (CBS, 2021).

3.1.2 Data Management

Access to data will be facilitated through the 'cbsodataR' module, which connects directly to CBS's OpenData API service, negating the need to store data locally and further reducing privacy risks. For the transformation and calculation of variables used in the model the packages ‘dyplr’ and ‘tidyr’ are employed. To join the different data sources together, the standardised regional code variable was used. Data visualisations and maps will be generated using R packages ‘ggplot2’ for graphs and ‘sf’ ('Simple Features') for mapped representations.

3.2 Operationalisation of Variables and Cases

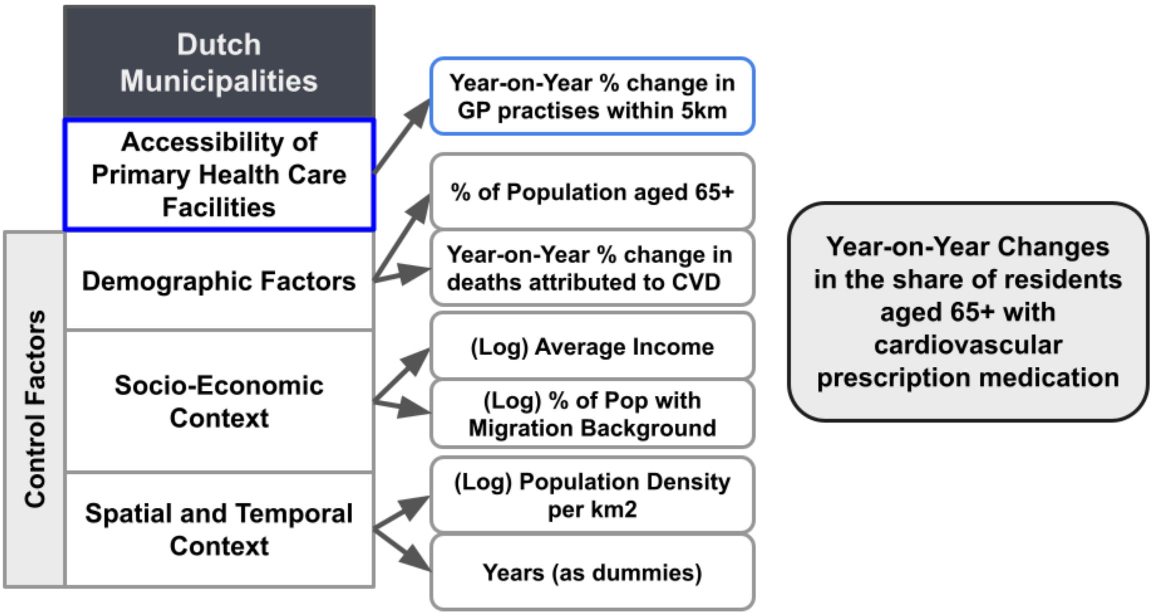


Figure 2: Concepts to Variable Operationalisation

From the sourced data, the concepts laid out in the conceptual framework are operationalised, as visualised in Figure 2. The main variable of interest for this study is highlighted in blue.

First, to investigate the possible effects of the loss of GP practices, the variable of ‘Rate of Population (aged 65+) with dispensed cardiovascular prescription medication (ATC code ‘C’) per municipality’ will be used as the health outcome measure. The year-on-year changes in the percentage of the population that has been prescribed cardiovascular medication can give insight into the changes in prescription patterns. As the normal processes of ageing lead to the degradation of the cardiovascular system, a decline in this rate in combination with increased distance to GP practices might indicate that the healthcare needs of the population are not being met.

The data does not include medication dispensed in nursing homes or hospitals, which is further useful to establish the possible relation between the two variables, as those in care facilities would not have the same accessibility challenges as those who have to either acquire medication themselves or rely on others to do so. Nonetheless, many other factors influence trends in prescription behaviour, which has to be considered for this analysis.

The medication group aimed at the cardiovascular system is further divided into 10 subgroups. However, with all medication under the ATC categorisation of C increasing in prevalence with higher ages, and with all relating to cardiovascular disease in different forms, all are included in the analysis. This category includes the most commonly prescribed types of cardiovascular medications: renin-angiotensin system agents (*C09*), antilipaemics (*C10*), and beta blockers (*C07*).

The measure of year-on-year changes in the spatial accessibility of GP practices within a municipality is the independent variable. This will be derived from the 'GP practices within 5km' variable, where a positive percentage change from one year to another indicates the gain of primary healthcare functions within 5km, averaged for residents of a municipality, and through that a negative change indicates a reduction. With the focus on older adults, this decline might be especially detrimental to the quality and use of care, as with rising age comes not only an increased need for general care, but also the additional obstacle of age-specific mobility challenges.

While it is challenging to include all municipal-level factors that have, or might have, an influence on the outcome variable of prescription rates, the dataset with key figures provides some general options for control factors. The first factor to be included in the analysis is the demographic composition of the municipality, specifically the share of the population that is aged 65 and above. A population with a higher than average share of elderly residents might experience the negative effects of healthcare function loss more severely, as a higher share of the population has higher needs for primary care. On the other hand, primary healthcare in areas with a high share might have more standardised approaches to meet the healthcare needs of local residents, and through that may mitigate negative effects.

Another interesting factor to consider in the analysis of the changes in the share of the elderly population with cardiovascular prescription, are the changes in mortality attributed to cardiovascular disease. For this, the variable 'number of deaths caused by diseases of the cardiovascular system per year per municipality' from the key figures dataset is used to calculate the year on year changes in the share of cardiovascular-caused deaths. Especially in municipalities where there is both a reduction of healthcare accessibility and a reduction in the rate of cardiovascular medication, it is necessary to make sure that this is not correlated to excess deaths attributed to cardiovascular diseases.

To account for the different social and economic contexts of each municipality, the variables 'average standardised income' and 'percentage of the population with migration background' are included in the model. For both, the natural log is taken to normalise the otherwise skewed distributions. However, despite this transformation, meant to account for the wide range of values, some municipalities still have very high outlier values of average income, that form the long tail of the normal distribution.

Acting as a control variable for the spatial context of any given municipality, is the population density per km^2, which is again also transformed with the natural logarithm. This is important to include, not only because the the residential context of an individual has an impact on their heart health, but further, areas with low population density are more rural, and tend to have an above average share of older residents, with a lower inherent density of facilities, as well as often experiencing higher rate of general function loss. Additionally, the reliance on personal vehicles is also often higher in these regions, which might increase resilience to function loss for the research population to a certain extent.

Lastly, the years of observations are included as factored dummy variables. This is to ensure that national changes in prescription behaviour are controlled for, and to give insight into possible overarching trends in regards to primary health and associated policy changes. While data for 2013 is available and included in the sourced data, due to the transformation into year-on-year percentage changes of multiple variables, the first year of data is missing values, as it can not be computed without a prior year of data.

There are many more factors that would be suitable to be included in the statistical model, such as markers of the social and economic well being of the residents. While other factors were considered, at this level of analysis, it is hard to avoid issues with multicollinearity between factors. The limited nature of this research as a bachelor's thesis calls for a narrowed scope, with an acknowledgment to the limitations in the validity of the findings. Further, as the proposed combinations of the main factors is a novel and untested approach, the need for further investigation into the relationship is evident.

3.3 Sample

The descriptive statistics, shown in Table 1, provides information on general variables to describe the researched sample of municipalities, alongside the transformed and untransformed factors used for the analysis. Key insights include the population size, with a mean of 47,894, varying significantly across municipalities as indicated by a high standard deviation (SD) of 72,300 and a range from 3,578 to 862,965. The data reflects a predominantly rural composition with 65% of the 334 unique municipalities classified as rural.

The percentage of the population aged 65+ with cardiovascular medication averages at 63.81%, and shows a relatively consistent distribution with an SD of 2.98. Finally, the year-on-year percentage change in GP practices within 5km presents a slight decline with an average of -0.28%, underscored by the variability in this change, as the SD of 0.48 and the range from -2.26% to 2.10% suggest.

The study utilises a balanced panel dataset spanning 6 years, from 2013, the first year with complete data, to 2019, chosen to exclude the distortions of the Covid-19 pandemic. Though changes in municipal boundaries over time (CBS, 2019) due to absorption or combination of smaller municipalities led to the exclusion of municipalities that did not exist for the entirety of the researched timeframe.

Table 1: Descriptive Statistics of the Final Sample

Variable	*N*	*Mean*	*SD*	*Median*	*Min*	*Max*
Population Total	2004	47894	72300	28949	3578	862965
Sex Ratio (M/F)	2004	0.99	0.03	0.99	0.88	1.15
Predominantly Rural (binary, 1 = rural) Rural: 223 unique, 1294 total Urban: 121 unique, 710 total	2004	0.65	0.48			
Population aged 65+ with Cardiovascular Medication (%)	2004	63.81	2.98	64.02	52.64	73.79
Year-on-Year % Change	2004	-0.28	0.48	-0.28	-2.26	2.10
GP practices within 5km	2004	9.62	10.40	6.10	1.00	87.00
Year-on-Year % Change	2004	-0.29	6.38	0	-53.85	50.00
Population aged 65+ (%)	2004	20.12	3.30	20.15	8.45	31.94
Deaths attributed to CVD (%)	2004	26.08	3.62	26.00	12.82	50.00
Year-on-Year % Change	2004	-0.39	4.44	-0.43	-22.62	20.46
% of Population with Migration Background	2004	15.42	8.46	13.50	3.10	54.60
Natural Log	2004	2.60	0.52	2.60	1.13	4.00
Average Standardised Household Income (yearly, in 1000 euro)	2004	30.44	4.08	29.80	22.80	71.50
Natural Log	2004	3.41	0.12	3.39	3.13	4.27
Population density (per km2)	2004	905.33	1044.71	483.00	57.00	6523.00
Natural Log	2004	6.28	1.02	6.18	4.04	8.78

Additionally, while a great effort was made to include all municipalities that existed for the research duration, the three least populated municipalities had too many extreme outlier values in multiple variables, even after transformations were applied. The islands Vlieland and Schiermonnikoog have extreme changes in the prescription rate for medications, with some years having the largest reduction, and the remaining the largest increases. This is likely due to its low population number and low population density. Schiermonnikoog also has extreme outliers, at both ends, for yearly change in GPs within 5km.

Lastly, Rozendaal, which is the least densely populated municipality on the Dutch mainland, isn't only among the municipalities with the highest average income, but also has the lowest share of the population aged 65+ with prescription medication. All of them have at least some extreme outliers in each of the variables transformed into year-on-year % changes, much more so than any other municipality, leading to the decision to exclude them from the final sample.

3.4 Analytical Strategy

To answer the research question "To what extent does the decline in local primary healthcare provisions influence cardiovascular medication prescriptions in municipalities in the Netherlands?" a pooled Ordinary Least Squares (OLS) regression analysis model based on Woolridge (2010) is employed. For this method, the cross-sectional multi-year data is 'pooled', which flattens the characteristics of the individual groups (municipalities in this case), treating each observation as if it came from the same group, ignoring any group-specific effects. The Null Hypothesis (H0) for this analysis is: "There is no relationship between the decline in local GP practices and the rates of cardiovascular prescriptions among the elderly in Dutch municipalities, when controlling for socioeconomic and age demographic factors".

The general mathematical formula for a pooled OLS is:

$$y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \epsilon_{it}.$$

In this formula, y_{it} represents the dependent variable y, which, for this study, is the year-on-year percentage change in rates of cardiovascular medication prescriptions of the population aged 65+ for municipality i in year t. The X_{kit} terms represent the independent variables that are hypothesised to influence the dependent variable. These include the factors as were determined in the previous section, for each municipality i in each year t. β_0 is the intercept of the model, whereas β_k are the coefficients that measure the impact of each independent variable on the dependent variable, and ϵ_{it} represents the error term, capturing all other factors affecting the dependent variable that are not included in the model. From this, the model is specified using the plm function in R as follows:

```
plm(
  formula = perc_change_Pop_65_up_Meds ~
            perc_change_GP_within_5km + perc_Pop_65_up +
            perc_change_deaths_CVD +
            ln_perc_Pop_migration_background +
            ln_avg_Standardized_Income + ln_pop_density_km2 +
            year,
  data = df_panel,
  model = "pooling",
  index = c("municipality", "year")
)
```

The formula argument in the function denotes the relationship being examined, the dependent variable, influenced by a set of predictors along with year dummies to control for temporal effects. The data parameter is set to df_panel, which contains our panel dataset, organised by municipalities over different years. The model parameter is explicitly set to "pooling", indicating that a pooled regression model is used, treating the data as a large cross-section rather than exploiting the panel structure. Lastly, the index parameter comprising "municipality" and "year" specifies the two-dimensional structure of our panel data, highlighting the repeated observations for each municipality over multiple years.

To more deeply explore the associations, other model specifications are also included and analysed. For a baseline understanding of the association between the variables ‘Year-on-Year % Change in Population aged 65+ with Cardiovascular Medication’ and the ‘Year-on-Year % Change in GPs within 5km ’ is established through the Null Model. Followed by integrating temporal dynamics with the Null Model + Years. Then, to address the first secondary research question, the variable ‘% of Population aged 65+’ is introduced, to analyse the influence of age demographic factors on the relationship. With regard to the second secondary research question, the full model as described above is stratified into a rural and urban model, to compare how the associations differ across the different spatial contexts.

For the coefficient estimators to be consistent, Error Terms need to be normally distributed and independent, with no presence of heteroskedasticity, and no serial autocorrelation. To address the requirement for serial independence, lagged variables in the form of Year-on-Year percentage change transformations are used for indicators for which that makes conceptual sense (see ‘Operationalisation of Variables and Cases’) and a quadratic measure of time is included in the model. To test these assumptions, the residuals are investigated and the Breusch-Pagan test for heteroskedasticity (Breusch & Pagan, 1978) and the Breusch–Godfrey test for serial correlation of the idiosyncratic parts of the error terms (Breusch, 1978; Godfrey, 1978) are used.

As complete homoscedasticity is hard to achieve with this kind of model design, a Clustered Robust Standard Error adjustment of type HC1 based on MacKinnon and White (1985), with the Arellano method to account for serial correlation (Arellano, 2003) is used. In the regression analysis, it is crucial to ensure that the estimated standard errors are robust to potential issues in the data, to address the inference of persistent heteroskedasticity and and within-group correlation. The choice of the Arellano method is based on its effectiveness in correcting for both within-group correlation and heteroskedasticity, offering a more reliable estimation of standard errors. Specifically, the clustering at the municipal level aims to address that observations within the same municipality are likely not independent and may share common unobserved characteristics (Cameron & Miller, 2015). The "HC1" specification is chosen due to the presence of heteroskedasticity and the relatively large number of observations (N) in the dataset. In cases of large samples, the HC1 type is preferable as it does not require a small-sample adjustment, while being robust to heteroskedasticity. (MacKinnon & White, 1985).

When dealing with spatial units, such as municipalities in this study, another form of autocorrelation that is important to consider is that of spatial autocorrelation. Considering Tobler’s first law of Geography (Tobler, 1970), it is clear that municipalities that are closer to each other will be dependent on each other, which breaks the assumption of independence. As Woolridge states (2010), this is a difficult problem to address, and is therefore often ignored in practical applications of pooled OLS models. While many methods have been developed, in the scope of this study, the issues of the model specification and the resulting limitations in the findings are acknowledged in the discussion.

4. Results and Analysis

4.1 Results of the pooled OLS

In this section, the results of the OLS pooled models are presented in Table 2, highlighting the significance of different variables between the different model specifications.

Table 2: Model Output

Year-on-Year % Change in Population aged 65+ with Cardiovascular Medication						
Variable	**Null Model**	**Null Model + year**	**Model Age**	**Full Model**	**Model Urban**	**Model Rural**
(Intercept)	-0.2758*** (0.0118)	-0.3398*** (0.0288)	-0.4120*** (0.0773)	-1.3475*** (0.3873)	-1.5142* (0.5933)	-1.3227** (0.4941)
Year-on-Year % Change in GPs within 5km	0.0053** (0.0017)	0.0029 (0.0016)	0.0029 (0.0016)	0.0032* (0.0016)	0.0007 (0.0028)	0.0040* (0.0019)
% of Population aged 65+			0.0038 (0.0039)	0.0062 (0.0044)	0.0155* (0.0070)	0.0036 (0.0056)
Year-on-Year % Change in CVD Deaths				0.0015 (0.0029)	0.0063 (0.0047)	0.0005 (0.0033)
Log % of Population with Migration Background				0.0106 (0.0368)	0.0866 (0.0559)	-0.0200 (0.0512)
Log Average Standardised Income				0.1968 (0.1215)	0.1997 (0.1651)	0.2451 (0.1730)
Population Density *(per km2)*				0.0322 (0.0180)	-0.0028 (0.0300)	0.0218 (0.0277)
Year *(ref: 2014)* *2015*		0.1011** (0.0389)	0.0989* (0.0391)	0.1002* (0.0394)	0.1170* (0.0504)	0.0937 (0.0540)
2016		0.2382*** (0.0365)	0.2341*** (0.0364)	0.2252*** (0.0362)	0.2793*** (0.0512)	0.1965*** (0.0489)
2017		0.1245*** (0.0372)	0.1190** (0.0381)	0.1029** (0.0389)	0.1187* (0.0558)	0.0934 (0.0509)
2018		-0.0238 (0.0378)	-0.0309 (0.0385)	-0.0499 (0.0394)	-0.0266 (0.0504)	-0.0657 (0.0543)
2019		-0.0606 (0.0368)	-0.0693 (0.0386)	-0.1054* (0.0412)	-0.0918 (0.0504)	-0.1212* (0.0563)
R^2 *(Adjusted* R^2*)*	0.0050 (0.0045)	0.0487 (0.0459)	0.0051 (0.0041)	0.0570 (0.0518)	0.1001 (0.0859)	0.0448 (0.0366)

Clustered Robust Standard Errors in Parentheses; Significance codes: *** p<0.001, ** p<0.01, * p<0.05

Initially, the 'Null Model' coefficients undergo notable changes when the temporal context is added, transitioning to the 'Null Model + Year'. In the Null Model, the factor for 'Year-on-Year

% Change in GPs within 5km' shows a positive coefficient at 0.0053 with a significant p-value below 0.01. With the addition of the year dummies, the coefficient decreases to 0.0029 and is no longer significant. The years 2015-2017 show a small positive effect (coefficients: 0.1011; 0.2382; 0.1245), with 2015 significant at p<0.01, 2016 and 2017 highly significant at p<0.001. For the remaining years the effects are slightly negative and not statistically significant. This indicates a non-linear temporal relationship of the factors. The explanatory value of the Null model is very low (R-Squared: 0.005; Adjusted R-Squared: 0.045), but increases nearly 10-fold to 0.0487 and 0.0459 when adding the temporal factors.

Further inclusion of '% of Population aged 65+' in 'Model Age' does not impact the relationship between the factors much, though the R-Squared values decrease again. The coefficient for 'Year-on-Year % Change in GPs within 5km' remains stable at 0.0029 (not significant), indicating a consistent, yet not statistically significant, relationship. The newly included '% of Population aged 65+' has a positive, yet non-significant effect (0.0038). The temporal effects remain largely consistent as well.

The 'Full Model' includes all control factors, as explained in the methodology section, and accounts for approximately 5.7% of the variance in the dependent variable (R-Squared: 0.057; Adjusted R-Squared: 0.0518). This suggests that while the included variables are relevant, there are also missing factors which play a significant role.
The percentage change in GP practices within 5km radius is statistically significant with a coefficient of 0.0032 (p-value: 0.0477) suggests that a 1% increase in GP's within 5km is associated with a 0.3% increase in prescription rates for cardiovascular medication. Further, the positive effect of '% of Population aged 65+' increases (0.0062), yet remains not significant. While none of the other control factors meet the standard level of significance of below 0.05, both the log of the average standardised income (coefficient: 0.1968) and the log of population density per km^2 (coefficient: 0.0322) showed positive relationships with the dependent variable. The temporal effects remain stable when compared to 'Model Age', though the coefficient of 2019 decreases and becomes significant at (coefficient: -0.1054, p<0.01).

Lastly, comparing the stratified models, 'Model Rural' and 'Model Urban', gives insight into how the association of factors differ between different spatial contexts. Both models are significant, though their R-Squared and Adjusted R-Squared differ, with the Rural model explaining less of the variance (R2: 0.0448, Adj. R2: 0.0366) compared to the Urban model (R2: 0.1001, Adj. R2: 0.0859). While the 'Year-on-Year % Change in GPs within 5km' coefficient is small and positively significant in the predominantly rural municipalities (0.0040, p<0.05), it is greatly reduced and not significant in 'Model Urban' (coefficient: 0.0007). Interestingly, the '% of Population aged 65+' presents a significant positive coefficient in 'Model Urban' (0.0155, p<0.05), but remains non-significant in 'Model Rural' (0.0036).
While the directions and strength of the year effects and the very high significance of 2016 remain largely stable, the temporal effects diverge. The positive coefficients of 2015 and 2017 are not significant in the rural model, and the negative coefficient of 2019 becomes insignificant in the urban model.

4.2 Discussion

A significant relationship between changes in local general practitioner availability and medication prescriptions highlights an effect of the spatial accessibility of primary healthcare in managing population cardiovascular health, at the scale of a 1% increase in the proximity of general practitioners leading to a 0.3% increase in cardiovascular medication prescriptions. This does support the first hypothesis, and answers the main research question. However the presence of spatial autocorrelation and potential serial correlation, alongside the low R-squared value, suggests that the model may be missing key explanatory variables. Alternative modelling approaches that better take the spatial and temporal dimensions into account, would be better suited to capture the complexity of these relationships.

4.2.1 Rural vs Urban differences

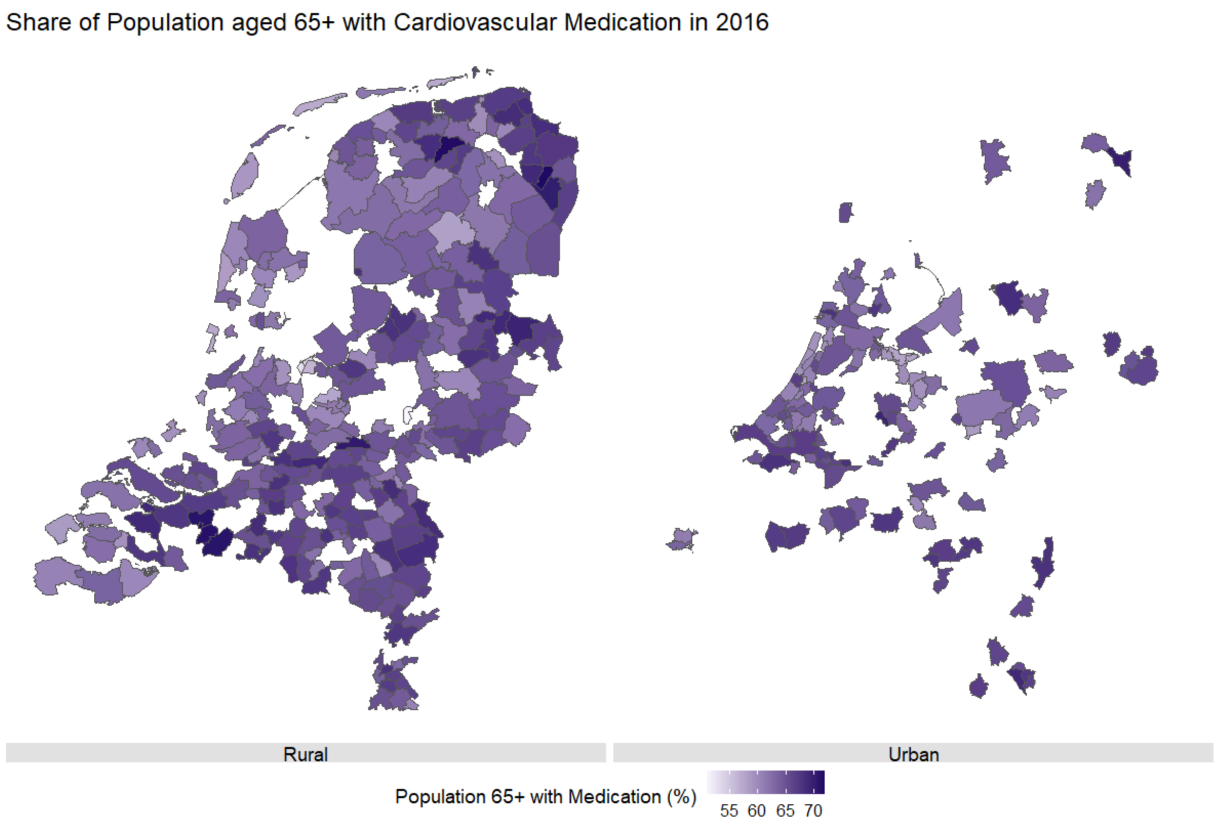


Figure 3: Rural and Urban Municipality Comparison Map

Aligning with Hypothesis 2, that the spatial context has an effect on the associations of the factors relating to GP accessibility and medication rates, the Year-on-Year % change in GP practises within 5km is only significant in the Rural model ($p<0.05$) and the share of population aged 65 and older was only significant in the Urban model. Additionally, the temporal effects diverge in between the two stratified models.

While the effect is very small, a 1% increase in GPs within 5km correlated to a 0.4% increase in prescription rates, this nonetheless aligns with the findings of Nemet and Bailey (2000), where rural residents were impacted by declining healthcare service presence, as utilisation decreased with higher distance from healthcare facilities.

It is important to note, that while the share of the population above age 65 was not a significant factor in the rural model, rural municipalities, on average, have a higher share of elderly residents. This also includes the primary healthcare providers themselves: they are home to a disproportionate share of GPs aged 50 and above (CBS, 2020a). With the fact that nationally, 15% of GPs are expected to retire in the next 5 years, and the already existing struggle to fill empty practices, this predicted decline has the potential to not only exacerbates spatial inequalities in healthcare access and put pressure on the quality of care (ABN Amro, 2023), but therefore might also directly influence prescription patterns for medications and health outcomes.

4.2.2 Age Demographics

Hypothesis 3 was that the proportion of the population over the age of 65 in a municipality has an effect on the prescription rates of cardiovascular medication. The findings show this to only be true in urban municipalities, as it was not a significant factor in the 'Model Age', nor in the full or rural model. Only in urban regions an increased proportion of elderly residents leads to higher rates of cardiovascular prescriptions.

While there are a wide possibility of factors that may explain this, one hypothesis is that within urban regions where there is less of a decline in services, municipalities are better at providing services based on their residents' demographic makeup (Goh et al., 2018; Gilliand et al., 2019). As such, an urban municipality with a higher share of elderly residents may already have measures and services in place to suit their residents' needs. The aforementioned studies also highlighted the importance of individual level factors, which are not well captured at this level of analysis. It seems logical that with lower proximity barriers due to the greater facility density and better transport options, individual barriers gain importance.

4.2.3 Temporal Factors

An additional consideration is the temporal impact on the model. The year of the data was added as a dummy variable, and highlighted 2016 as an unusual year. The plots shown in Appendix Figures 7.2.1-3, show the year-on-year changes in GPs within 5km, the year-on-year changes in the population over 65 with cardiovascular medication, and the share of the population over 65 for each municipality.

Overall, the proportion of the population over 65 is increasing every year. However, for both the change in GPs within 5km and the elderly population with cardiovascular medicine we see a peak in 2016. This does reflect a period where there was a change at the EU-level in the risk assessment criteria for cardiovascular medication (Piepoli et al., 2016). However, in 2015 the Dutch government introduced the Exceptional Medical Expenses Act (AWBZ), part of which decentralised the responsibility of elderly care to the municipalities (Government Netherlands, 2015).

Within all models that include the years as factors, 2016 is highly significant, but for the urban model 2015 and 2017 are also significant, unlike the rural model. This significance may represent the impact of the change introduced by AWBZ, which was more significant among the urban model where there is less unexplained variance. These, along with other factors that might be clearer to a local researcher, highlight that the results from each year are reflections of ongoing and continuous changes.

4.3 Discussion of Limitations

While the model is significant, it is important to acknowledge that the model choice is not the best fit for the data and specifications used. The pooled OLS model was chosen for its relative simplicity and interpretability, compared to other common econometric models used for panel data. Therefore, this model can not capture all the complexities, such as individual or time effects that fixed or random effects models could reveal. As mentioned previously, the year was significant in some cases, a sign of the likely impact of time effects.

In more specific detail, the Breusch-Pagan Lagrange Multiplier Test (Baltagi et al, 2003) indicated potential spatial autocorrelation and heteroskedasticity, though for the latter, the investigation of residual plots did not invoke concern (see: Appendix 7.2.1). Given the structure of the data used, in combination with the model choice, this is a common issue. To try and address this, a clustered robust standard errors adjustment, clustered on the 'group', is employed, as discussed in the Methodology section. While this is crucial for correcting the within-group correlation in panel data, this approach does not resolve the underlying issues in the model specification.

Another form of statistically problematic correlation to consider in this model, is that of spatial autocorrelation. The Moran's I test for spatial dependence was performed on all variables included in the model, suggesting that geographical factors influence the results. The pooled OLS model does not account for these spatial dependencies, and as such the assumption of independent error terms is violated, and this remains a limitation of this study. Further, there is a high possibility of omitted variables that could affect the results, which is a limitation inherent in regression analysis. Future research might consider additional variables that could influence the yearly changes in prescription rates in the older adult population.

5. Conclusion

While factors like smoking cessation and improved physical fitness have contributed to declining cardiovascular disease rates (Koopman et al., 2016), resulting in cancer replacing it as the leading cause of death in the Netherlands(CBS, 2020b), addressing cardiovascular health disparities remains critical. The findings of this study underscore the significant role of local healthcare infrastructure in cardiovascular medication prescriptions.

This research suggests that the changes in the availability of local primary healthcare practices have an effect on the cardiovascular prescription rates among the elderly in dutch municipalities. In rural regions, healthcare utilisation among the elderly may decrease as the proximity and availability of GPs decreases. As these rural regions are both elderly and seeing a trend in reduced services, measures to increase healthcare services are likely effective measures to improve cardiovascular health of the elderly in these regions. However, further research with more complex statistical models is needed to explore the additional factors impacting prescription rates in these regions.
Further, within urban municipalities we see a positive relationship between the proportion of elderly residents and cardiovascular prescription rates among the elderly, though the changes in spatial proximity are not significant in the more service dense contexts.

However, the complexity of healthcare utilisation and medication use, combined with significant geographical and temporal factors, warrants further investigation using more comprehensive models or methodologies to gain a more nuanced understanding of these dynamics. As such, this study and its findings are specific to the models and data used, and may not be generalisable to other settings or populations.

Acknowledgements

First and foremost, thank you to Liam for the loving support throughout the writing of this thesis. Thank you to Bailey, Gismo, and Ezra for sharing the warmth of my laptop with me, and fulfilling their duties as cat-shaped fidget toys. I also need to thank John Flournoy and Brayton Noll for their enthusiastic statistical insights. Last but not least, thank you to my supervisors, especially my primary supervisor Lara Bister - you write the best feedback ever.

6. References

6.1 Literature

Aerts, R., Vanlessen, N., Dujardin, S., Nemery, B., Van Nieuwenhuyse, A., Bauwelinck, M., Casas, L., Demoury, C., Plusquin, M., & Nawrot, T. S. (2022). Residential green space and mental health-related prescription medication sales: An ecological study in Belgium. Environmental Research, 211, 113056. https://doi.org/10.1016/j.envres.2022.113056

ABN Amro. (2023, May 4). *Huisarts? Bekijk de trends van 2023 in uw vakgebied.* (Online Report). Retrieved September 15, 2023, from: https://www.abnamro.nl/nl/prive/speciaal-voor/medici/tips-voor-medici/trends-huisartsenzorg.html

AHA (2023). Hearts and bodies change with age, heart disease treatments may need to change, too. American Heart Association (Online Article.) Retrieved 25th of September, 2023, from: https://newsroom.heart.org/news/hearts-and-bodies-change-with-age-heart-disease-treatments-may-need-to-change-too

Arellano, M. (2003). Panel data econometrics. OUP Oxford.

Baltagi, B. H., Song, S. H., & Koh, W. (2003). Testing panel data regression models with spatial error correlation. Journal of econometrics, 117(1), 123-150. https://doi.org/10.1016/S0304-4076(03)00120-9

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, 1287-1294. https://doi.org/10.2307/1911963

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian economic papers*, 17(31), 334-355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources*, 50(2), 317-372. https://doi.org/10.3368/jhr.50.2.317

CBS. (2019). Gemeentelijke indeling op 1 januari 2019. *Centraal Bureau voor de Statistiek Nederlands* (Online Article). Retrieved 27th of September, 2023, from: https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke-indelingen-per-jaar/indeling-per-jaar/gemeentelijke-indeling-op-1-januari-2019

CBS. (2020a). Huisarts vaker vrouw en gemiddeld jonger. *Centraal Bureau voor de Statistiek Nederlands* (Online Report). Retrieved September 27, 2023, from: https://www.cbs.nl/nl-nl/nieuws/2020/28/huisarts-vaker-vrouw-en-gemiddeld-jonger

CBS. (2020b). *54 percent of deaths in 2019 caused by cancer or CVD. Centraal Bureau voor de Statistiek Nederlands* (Online Report). Retrieved September 15, 2023, from: https://www.cbs.nl/en-gb/news/2020/27/54-percent-of-deaths-in-2019-caused-by-cancer-or-cvd

CBS. (2021, April 19). Privacy. *Centraal Bureau voor de Statistiek Nederlands* (Webpage). Retrieved January 9, 2024, from https://www.cbs.nl/en-gb/about-us/who-we-are/our-organisation/privacy

CBS. (2023). Elderly People. *Centraal Bureau voor de Statistiek Nederlands* (Online Report). Retrieved September 24, 2023, from: https://www.cbs.nl/en-gb/visualisations/dashboard-population/age/elderly-people

Gilliland, J. A., Shah, T. I., Clark, A., Sibbald, S., & Seabrook, J. A. (2019). A geospatial approach to understanding inequalities in accessibility to primary care among vulnerable populations. *PloS one*, *14*(1), e0210113. https://doi.org/10.1371/journal.pone.0210113

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica: Journal of the Econometric Society,* 1293-1301. https://doi.org/10.2307/1913829

Goh, C. E., Mooney, S. J., Siscovick, D. S., Lemaitre, R. N., Hurvitz, P., Sotoodehnia, N., Kaufman, T. K., Zulaika, G., & Lovasi, G. S. (2018). Medical facilities in the neighborhood and incidence of sudden cardiac arrest. *Resuscitation*, 130, 118-123. https://doi.org/10.1016/j.resuscitation.2018.07.005

Government Netherlands (2015) Decentralisation of government tasks. *Government.nl* (Online Report). Retrieved December 13, 2023, from: https://www.government.nl/topics/municipalities/decentralisation-of-government-tasks

Koopman, C., Vaartjes, I., Blokstra, A., Verschuren, W. M. M., Visser, M., Deeg, D. J. H., Bots, M. L., & Van Dis, I. (2016). Trends in risk factors for coronary heart disease in the Netherlands. *BMC Public Health, 16*(1), 835. https://doi.org/10.1186/s12889-016-3526-7

Luengo-Fernandez, R., Walli-Attaei, M., Gray, A., Torbica, A., Maggioni, A., Huculeci, R., Bairami, F., Aboyans, V., Timmis, A., Vardas, P., Leal, J. (2023). Economic burden of cardiovascular diseases in the European Union: a population-based cost study. *European Heart Journal.* 44. 10.1093/eurheartj/ehad583

Nemet, G. F., & Bailey, A. J. (2000). Distance and health care utilization among the rural elderly. *Social Science & Medicine*. https://doi.org/10.1016/s0277-9536(99)00365-2

NHG (2019, June). Cardiovasculair risicomanagement. *NHG-Standaard*, M84. Retrieved January 21, 2023, from https://richtlijnen.nhg.org/standaarden/cardiovasculair-risicomanagement#volledige-tekst-richtlijnen-diagnostiek

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of econometrics*, 29(3), 305-325. https://doi.org/10.1016/0304-4076(85)90158-7

McDougall, C. W., Hanley, N., Quilliam, R. S., Bartie, P. J., Robertson, T., Griffiths, M., & Oliver, D. M. (2021). Neighbourhood blue space and mental health: A nationwide ecological study of antidepressant medication prescribed to older adults. *Landscape and Urban Planning*, 214. https://doi.org/10.1016/j.landurbplan.2021.104132

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17-23. https://doi.org/10.2307/2332142

Padeiro, M. (2018). Geographical accessibility to community pharmacies by the elderly in metropolitan Lisbon. *Research in Social and Administrative Pharmacy*, 14(7), 653-662. https://doi.org/10.1016/j.sapharm.2017.07.014

Piepoli, M. F., Hoes, A. W., Agewall, S., Albus, C., Brotons, C., Catapano, A. L., Cooney, M. T., Corrà, U., Cosyns, B., Deaton, C., Graham, I., Hall, M. S., Hobbs, F. D. R., Løchen, M. L., Löllgen, H., Marques-Vidal, P., Perk, J., Prescott, E., Redon, J., Richter, D. J., … ESC Scientific Document Group (2016). 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *European heart journal*, 37(29), 2315–2381. https://doi.org/10.1093/eurheartj/ehw106

RIVM (2023, December 9). RIVM identifies three major challenges for public health and healthcare in period to 2050. *National Institute for Public Health and the Environment* (Online Report). Retrieved January 21, 2024 from: https://www.rivm.nl/en/news/rivm-identifies-three-major-challenges-for-public-health-and-healthcare-in-period-to-2050

Rodgers, J. L., Jones, J., Bolleddu, S. I., Vanthenapalli, S., Rodgers, L. E., Shah, K., Karia, K., & Panguluri, S. K. (2019). Cardiovascular Risks Associated with Gender and Aging.

Journal of cardiovascular development and disease, 6(2), 19. https://doi.org/10.3390/jcdd6020019

Rushworth, G. F., Cunningham, S., Pfleger, S., Hall, J., & Stewart, D. (2018). A cross-sectional survey of the access of older people in the Scottish Highlands to general medical practices, community pharmacies and prescription medicines. *Research in Social and Administrative Pharmacy*, 14(1), 76-85. https://doi.org/10.1016/j.sapharm.2017.01.002

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1), 234-240. https://doi.org/10.2307/143141

van Vulpen, B., Bock, B. B., & van den Berg, C. F. (2023). Does regional decline trigger discontent? Unravelling regional development in the Netherlands. *Regional Studies,* 1-14. https://doi.org/10.1080/00343404.2023.2231031

WHO (2023) C CARDIOVASCULAR SYSTEM. *WHO Collaborating Centre for Drug Statistics Methodology* (Website). Retrieved January 22, 2024: https://www.whocc.no/atc_ddd_index/?code=C

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. *MIT press*. http://www.jstor.org/stable/j.ctt5hhcfr

6.2 Data Sources

CBS StatLine [2013-2019a] Personen met verstrekte geneesmiddelen; regio (gemeente). Dataset: 83251NED https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83251NED/table

CBS StatLine [2013-2019b] Proximity to amenities; distance location, regional figures. Dataset: 80305ENG https://opendata.cbs.nl/statline/#/CBS/en/dataset/80305ENG/table?ts

CBS StatLine [2013-2019c] Regionale kerncijfers Nederland. Dataset: 70072NED https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072ned/table

CBS StatLine [2013-2019d] Maatstaven Financiële-verhoudingswet (Fvw), regio, 2007 - 2023.
Dataset: 60039FVW https://opendata.cbs.nl/statline/#/CBS/nl/dataset/60039fvw/table

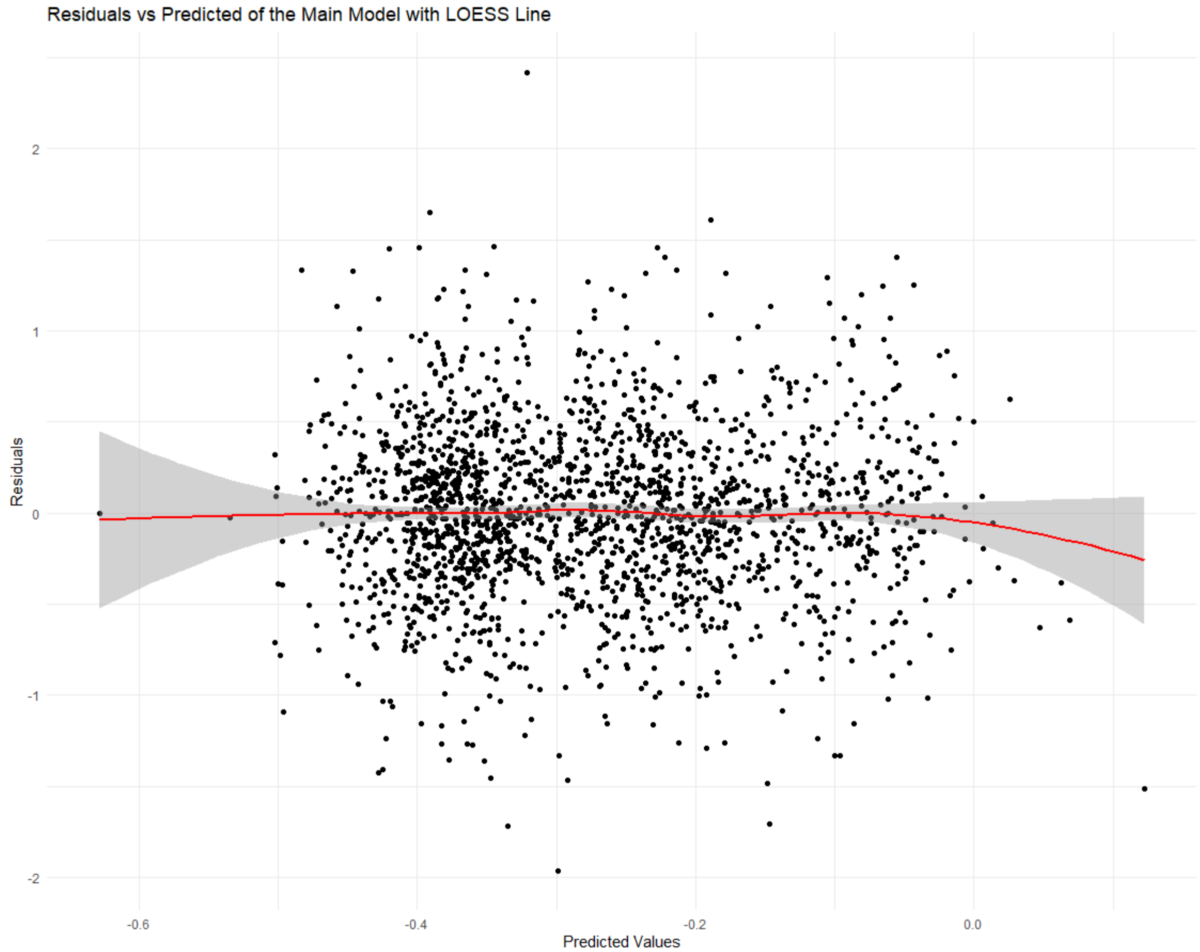
PDOK & CBS [2014-2019] Dataset: CBS Gebiedsindelingen https://www.pdok.nl/introductie/-/article/cbs-gebiedsindelingen

6.3 Script Repository

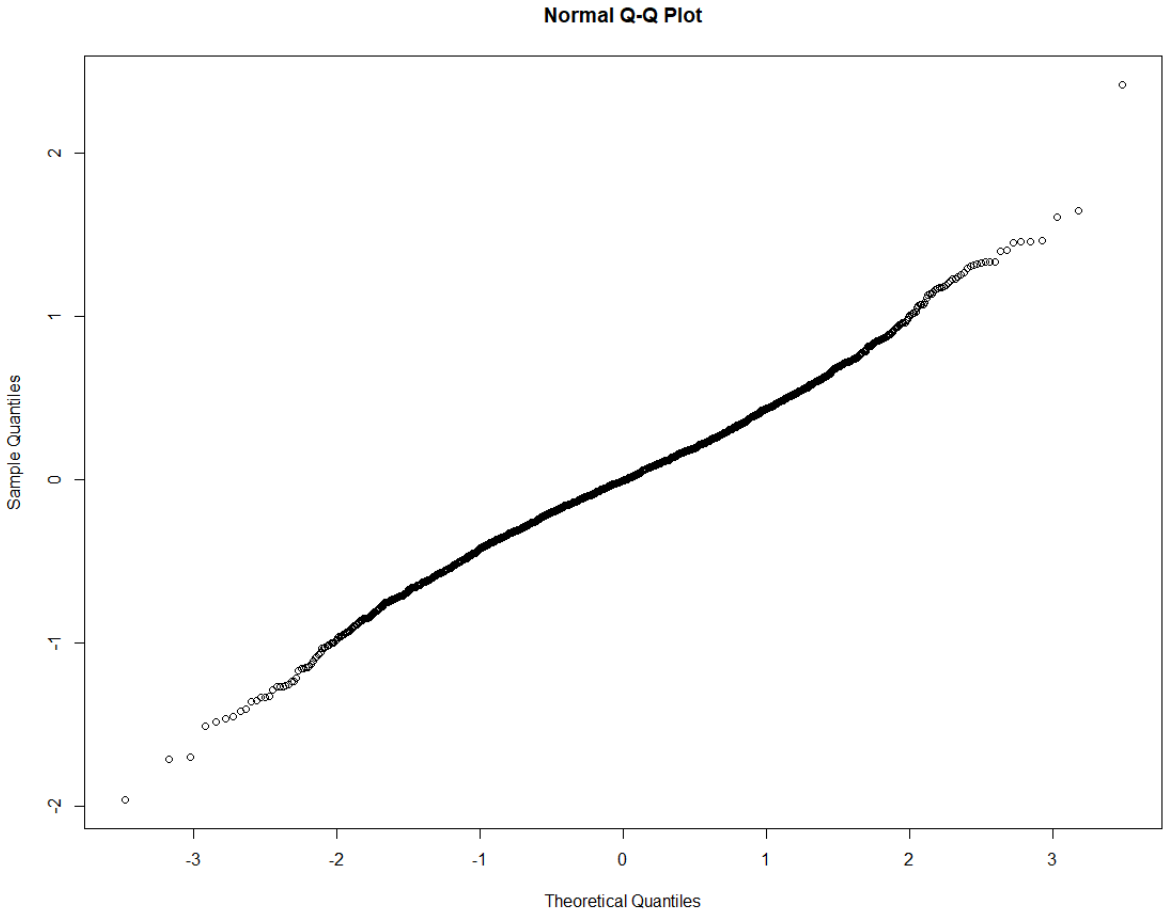
The R scripts and data used in this study are available on a public GitHub Repository: https://github.com/TinaTiresome/spatial-heart-analysis

7. Appendix

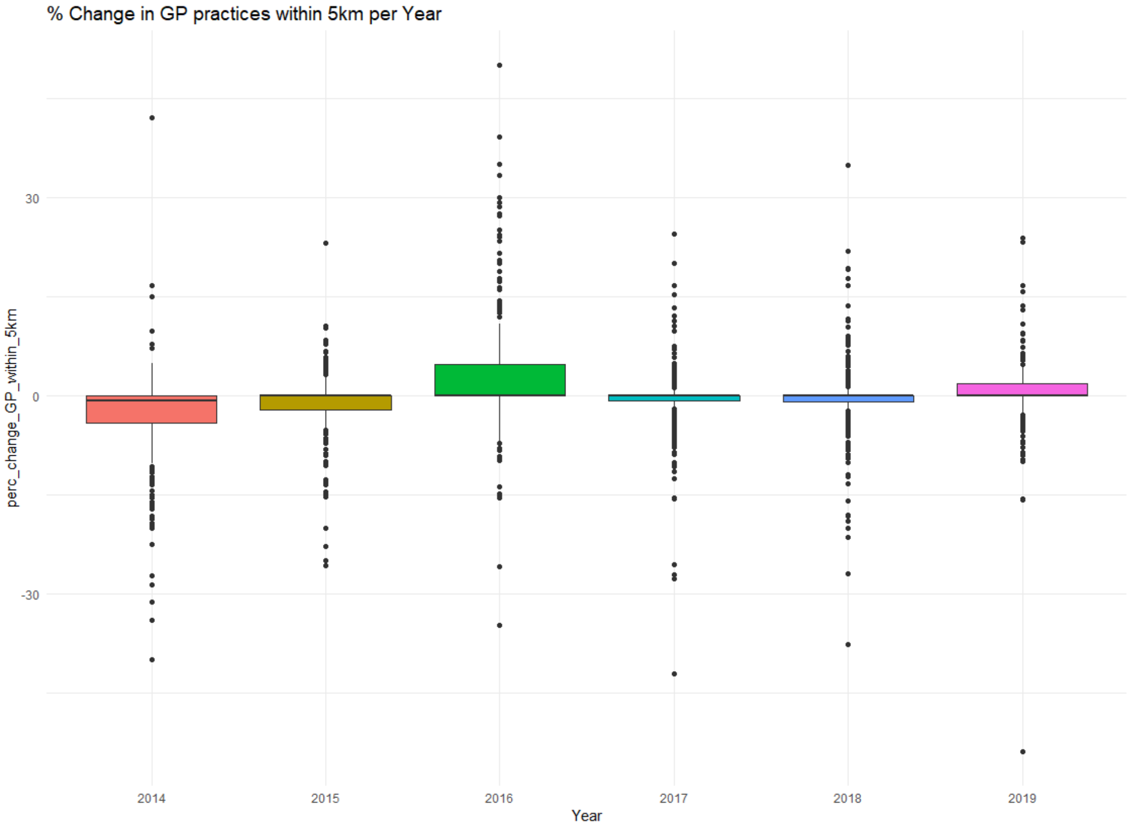
7.1.1 Residual Plot



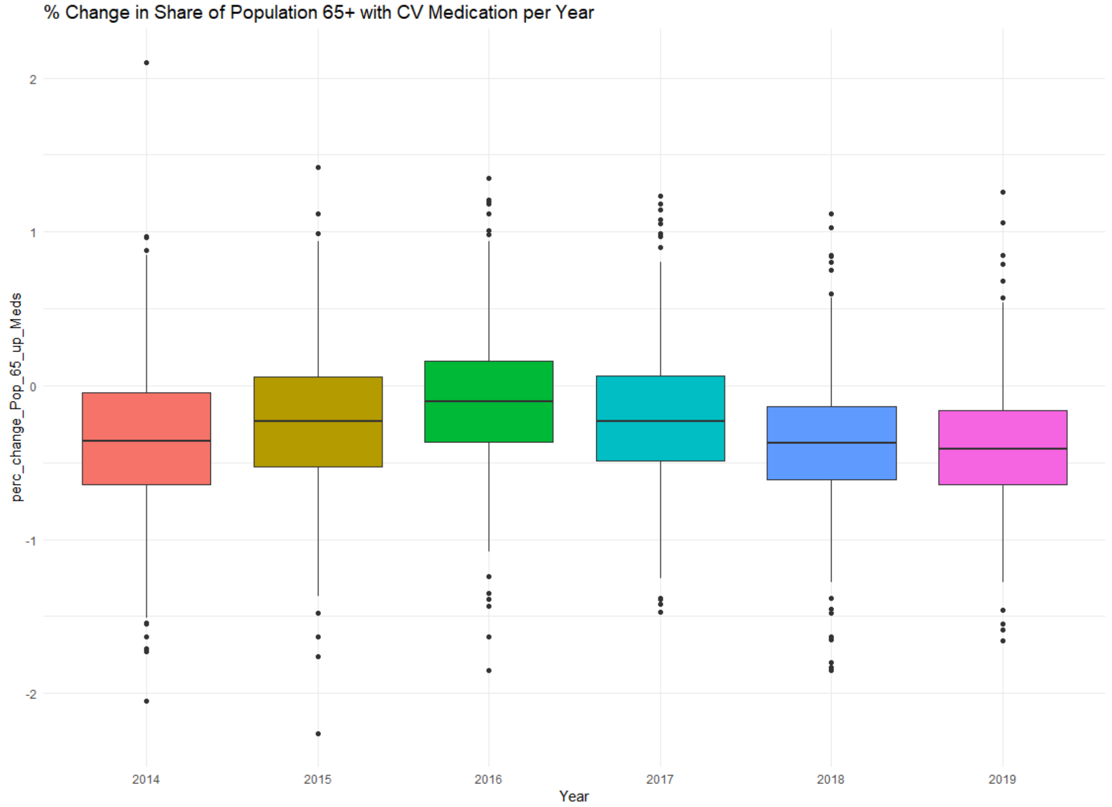
7.1.2 Residual Q-Q Plot for Normality of Error Term Assumption



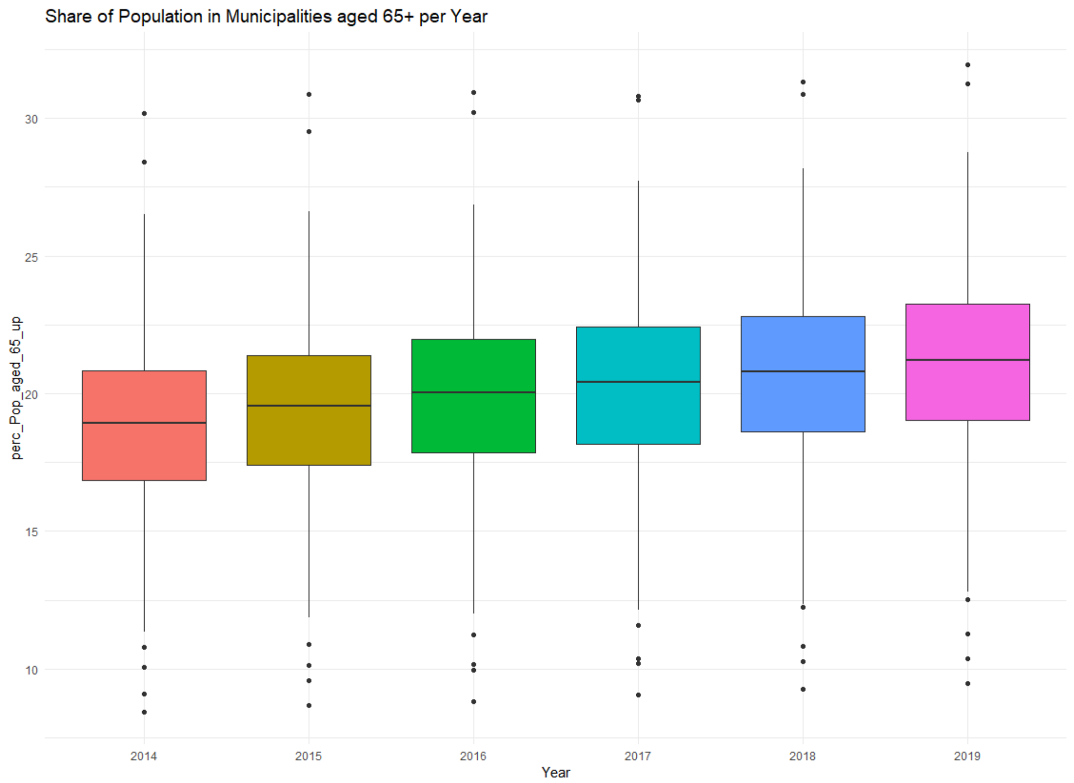
7.2.1 Boxplot Year on Year Changes in GPs within 5km per year



7.2.2 Boxplot Year on Year Changes in Population 65+ with CV medication per year



7.2.3 Boxplot Share of Population aged 65+ per year



7.3 Data Management Plan

1. General	
1.1. Name & title of thesis	*Impact of declining GP practices on dispensed Prescriptions of Cardiovascular Medications in Elderly Populations of Dutch Municipalities* Christina Bollmann - S4424743
1.2 (if applicable) Organisation. Provide details on the organisation where the research takes place if this applies	Not applicable

2. Data Collection - the creation of data	
2.1. Which data formats or which sources are used in the project? For example: - theoretical research, using literature and publicly available resources - Survey Data - Field Data - Interviews	Publically available data from CBS Centraal Bureau voor de Statistiek through API. Municipality-level data.
2.2. Methods of data collection What method(s) do you use for the collection of data. (Tick all boxes that apply)	- Structured individual interviews - Semi-structured individual interviews - Structured group interviews - Semi-structured group interviews - Observations - Survey(s) - Experiment(s) in real life (interventions) - **Secondary analyses on existing data sets** (if so: please also fill in 2.3) **- Public sources (e.g. University Library)** - Other (explain):
2.3. (If applicable): if you have selected ‘Secondary analyses on existing datasets’: who provides the data set?	- Data is supplied by the University of Groningen. **- Data have been supplied by an external party. (Please mention the party here):** **CBS (Centraal Bureau voor de Statistiek Nederland)**

3. Storage, Sharing and Archiving	
3.1 Where will the (raw) data be stored	- X-drive of UG network

during research? If you want to store research data, it is good practice to ask yourself some questions: - How big is my dataset at the end of my Research? - Do I want to collaborate on the data? - How confidential is my data? - How do I make sure I do not lose my data?	- Y-drive of UG network - (Shared) UG Google Drive - Unishare **- Personal laptop or computer** - External devices (USB, harddisk, NAS) **- Other (explain):** The data is pulled from the data API service offered by CBS. Therefore, it is in theory not even necessary to store it on my personal machine at all, though I do save a copy to reduce data load times.
3.2 Where are you planning to store/archive the data after you have finished your research? Please explain where and for how long. Also explain who has access to these data NB do not use a personal UG network or google drive for archiving data!	- X-drive of UG network - Y-drive of UG network - (Shared) UG Google Drive - Unishare - In a repository (i.e. DataverseNL) **- Other (explain):** **The code with the code for the data import, data merging, variable transformation and data analysis will be hosted on a public GitHub.** The retention period will be **at least 3 years** years.
3.3 Sharing of data With whom will you be sharing data during your research?	**- University of Groningen** - Universities or other parties in Europe - Universities or other parties outside Europe - I will not be sharing data

4. Personal Data	
4.1 Collecting personal data Will you be collecting personal data? If you are conducting research with personal data you have to comply to the General Data Privacy Regulation (GDPR).	Yes / **<u>No</u>**