The importance of income inequality as a determinant for Covid-19 testing allocation and disease control policy

Bachelor Thesis– Policy Economics

Emma J. Koch (511288)
Erasmus school of Economics, Erasmus University Rotterdam
Supervisor: Dr. Josse Delfgaauw
July 30, 2021

The recent rise of Covid-19 has taken over the world. People are unequally affected by the disease and the many measures to contain the spread. This paper examines the correlation between average income and the Covid-19 infection rate for different areas in the Netherlands. First, an analysis is performed to estimate the relation between average income and the Covid-19 infection rate for 338 Dutch municipalities. Second, an analysis with 27 areas in Rotterdam-Rijnmond is used to estimate the relation between average income and the relative infection rate, the rate at which an area got tested, and the chance of testing positive if tested. Both analyses estimate a statistically significant negative correlation between average income and the relative infection rate. The second analysis shows a positive relation between average income and the rate at which an area was tested and a negative relation between average income and the chance a person tests positive on Covid-19. In addition, it shows that only focusing on the relative infection rate falls short of explaining differences in infection rates. The infection rate, testing rate and chance to test positive are unequally distributed among areas, at the expense of lower income areas. Policy makers should therefore focus largely on lower income areas, which is favorable from an equality and efficiency perspective.

1. Introduction

"Together, we'll get the coronavirus under control", is the anti-Covid-19 slogan of the government of the Netherlands. However, is there a "together" if communities are not equally sacrificing and having different risks of getting contaminated? In this paper the relationship between average income and the Covid-19 infection rate is examined for different areas in the Netherlands. In addition, the rate at which an area was tested and the chance a person tested positive on Covid-19, are seen as two of the mechanisms influencing the Covid-19 infection rate. The relationship between average income and these two mechanisms are tested in this paper for areas in Rotterdam-Rijnmond.

Established research has shown that major historic epidemics result in either two different mechanisms: redistribution towards the poor, or extermination of the poor (Alfani, 2020). For example, cholera and the black dead are seen as equalizers, whereas the plague and Spanish influenza have caused significant increases in inequalities (Alfani, 2020; Bambra et al., 2020). Nowadays, since Covid-19 has taken over the world, the inequalities set off by pandemics are becoming increasingly relevant. The first meeting of the European Health Security Committee was held on the 24th of January 2020. And ever since, many policy makers and researchers tried to grasp the relation between health, economic, social risks and Covid-19 (Alfani, 2020; Bambra et al., 2020). Recent literature finds that low income groups have a higher infection rate of Covid-19. In addition, the financial damage as result of contamination measures of Covid-19 is unequally distributed. In the following section the impact of contamination measures and inequalities in Covid-19 infections are discussed more elaborate.

This recent literature, exposing the income inequalities in relation to Covid-19, is in line with findings in the field of health economics. The relationship between health and economics is a broadly studied subject. Research suggest that a higher income improves an individual's health and better health improves someone's income (Weil, 2013). Therefore, the relation seems to run in both directions, which also seems the case with income and Covid-19

This paper analyses the relation between economic inequality and the Covid-19 infection rate. To investigate this relation relevant and highly recent data from areas in the Netherlands is used. Research on economic inequality and Covid-19 is mostly done in countries like Germany, America and England. This research helps to understand the relation in the Netherlands specific. Moreover, this paper adds to existing research because it uses relevant and highly recent data. While other research is often established at the beginning of the Covid-19 crisis, or focused on a smaller sample of people.

Secondly, by examining the relation between average income and the infection rate this research helps to explain differences in infection rates between areas in the Netherlands. The government of the Netherlands regularly reports contamination figures regarding infections per 1000,000 inhabitants of a neighborhood, municipality or GGD region. These present-day figures often fall short of an explanation about the differences in infection rates between the areas. Those insights are relevant for society, researchers and policy makers since good crisis management, like fighting a pandemic, requires tangible and clear insights.

In addition, the relation between average income is evaluated on the rate at which an area was tested and the chance a person tests positive on Covid-19. As a result, this paper helps to confirm that only evaluating the infection rate does not give a sufficient explanation of the spread of Covid-19. Firstly, only actually tested infections will come up in the Covid-19 infection rate. Therefore, the willingness to test for Covid-19 influences the Covid-19 infection rate. If people do not test for Covid-19 the infection rate will go down, but that does not mean the disease is not prevalent. Moreover, the differences in willingness to test can influence the risks of getting Covid-19. If people do not know they are infected, because they do not test, they go about their daily business while being contagious. There may even be Covid-19 outbreaks going on in areas unnoticed. These health risks will be undetected in the Covid-19 infection rate. Therefore, the rate at which an area was tested and the chance a person tested positive on Covid-19 seem to play an important role in the spread of Covid-19. Examining these two mechanisms helps understand the development of Covid-19 infection rates across the Netherlands. Especially, the differences in the rate at which an area was tested and the chance a person tested positive can help to understand the inequalities in Covid-19 infection rates and the spread of Covid-19. Once again, a good grasp of these influences is crucial in fighting a pandemic efficient and fairly.

Research on the determinates of testing incidence and Covid-19 infections analyzed these mechanisms in New York City (Borjas, 2020). However, this research dates from April 2020, the beginning of the pandemic. Yet, this research is focused on aggregated data from June 2020 until May 2021 and therefore the data is more developed and recent. Moreover, this paper is focused on the Netherlands.

Lastly, the Covid-19 crisis affects the whole world in different ways. For example, the National Institute for Public. Health and the Environment (RIVM) has reported a total of 17,775 deaths as result of Covid-19 in the Netherlands. (RIVM, n.d.). Besides, a lot of people have to stay in isolation or self-quarantine. One Covid-19 infection on average leads to quarantine for 2 household members, 2 close contacts and 3 to 9 other contacts. (RIVM, 2021). In addition, there are a lot of other ways the Covid-19 pandemic influences lives. Therefore, it is desirable to contain the spread as much as possible. This research will help achieve that social goal.

For this research ordinary linear squares regressions are used to estimate the relation between average income and multiple ratios. In this paper two analyses are distinguished: the municipality and GGD Rotterdam-Rijnmond analysis. The municipality analysis uses 338 municipalities in the Netherlands as observations for the regression and the GGD Rotterdam-Rijnmond analysis 27 areas in the GGD region Rotterdam-Rijnmond. For the municipality analysis average income is only tested on the relative infection rate of Covid-19. Besides the relative infection rate, the GGD Rotterdam-Rijnmond also tests average income on the rate at which an area got tested and the chance of testing positive if tested. For all regression equation important confounding variables are added to reduce the problem of omitted variable bias.

To perform previous explained regression equations, data is needed. For the municipality analysis data of the National Institute for Public Health and the Environment (RIVM) and Statistics Netherlands (CBS) is used (National Institute for Public Health, n.d.; Statistics Netherlands, n.d.) and for the GGD Rotterdam-Rijnmond analysis data of GGD Rotterdam-Rijnmond and Statistics Netherlands (CBS) is used (GGD Rotterdam-Rijnmond, n.d.; Statistics Netherlands, n.d.). Both the analysis use averages or percentages for the independent variables and cumulative amounts for the dependent variables. The cumulative amounts of positive tests and tests taken by the GGD are from June 1th 2020 onwards, since testing occurred very sporadic in the beginning of the pandemic and from June 1th onward everyone was able to test for free at GGD locations in the Netherlands (Rijksoverheid, n.d.). Area and municipality characteristics are therefore not likely to have affected the allocation of testing resources.

Both analyses find a statistically significant negative relation between average income and the relative infection rate. Therefore, a decrease in the average income of an area is expected to lead to an increase in the Covid-19 infection rate. Yet, the GGD Rotterdam-Rijnmond analysis shows that focusing only on the relative infection rate falls short of what is really going on with Covid-19 infections. The results show a positive and statistically significant relation between average income and the rate at which an area was tested and a statistically significant negative relation between average income and the chance a person tests positive on Covid-19. Therefore, residents in wealthier areas appear to have tested for Covid-19 at a higher rate and have lower chances of testing positive if tested than residents of poorer areas.

These findings could be due to a broad range of aspects, which are reviewed in the discussion. However, concluding, the results show that it is important from an efficiency and equality perspective to let policy makers focus on areas with relatively low incomes. Since the number of actual infections in these lower income areas will presumably be a lot higher than the reported number of infections, which will result in higher health risks for these areas. Moreover, the results can only be interpreted as correlations, because the results still have potential confounding factors which are not controlled for in the regressions.

The remainder of this paper is organized as follows. In Section 2 an overview of the related literature will be presented. Section 3 describes the data, methodology and descriptive statistics for the municipality and GGD Rotterdam-Rijnmond analysis. Section 4 examines the results of the two analyses. The results are discussed in Section 5 and Section 6 concludes this thesis research.

2. Related literature

This section provides an investigation of the existing literature relevant for the current research. Covid-19 is not the first pandemic in the world, a broad range of research has been developed over pandemics in history. Several studies of inequality trends in past major pandemics have tried to accentuate the role major pandemics played in leveling inequality (Alfani, 2020). The paper of Alfani (2020) covers preindustrial epidemics and shows that distributive outcomes depend on the institutional framework in place at the time of the epidemic. Moreover, it shows that major lethal epidemics resulted in either two different mechanisms: redistribution towards the poor, or extermination of the poor. For example, cholera caused improvement in health and living conditions for the whole population and the black dead is seen as a great equalizer. On the other hand, the plague increased mortality among the poorest and to add, recent studies have shown that there have been significant inequalities in the Spanish influenza pandemic of 1918 (Bambra et al., 2020). Moreover, the more recent HIV/AIDS epidemic demonstrated, and still demonstrates, the structural inequalities between different communities (Bowleg, 2020). Concluding, inequality trends in past pandemics proceeded very differently.

Before an investigation of the relation between economic inequality and Covid-19, the broader relation of health and economics will be explored. There is a large part of literature dedicated to the relationship between health and economics. Weil (2013) explains that the causality between the two runs in both directions; higher income improves individual's health. An individual with a higher income can for example buy higher quality food, housing, medicine and more. Vice versa, better health conditions improve income. A healthier nation can for example work and study more productively, given that being unhealthy, sick or dying costs time and effort.

An important note to make is that an improvement in the health status of a country or area, overall does not necessarily mean an improvement in health equality. In the 90's the general health status of Western European countries improved rapidly. However, during this period of growth, the inequalities in health risks grew (Goran et. al., 1991). Differences in health conditions and risks are a result of unequal exposure to social determinates of health, like financial resources (Bambra et. al., 2020).

Although past pandemics seem to lead to significantly different distributive outcomes, multiple studies lay out the channels through which Covid-19 pandemic and inequality seem related. Alfani (2020) mentions two ways Covid-19 might lead to increasing inequality in society. First, by a higher infection rate of Covid-19 for the poor, higher health risks. Second, by more damage, financial, damage as result of the measures to reduce the spread of Covid-19, for the poor. These two ways seem consistent with other research that tries to expose the ways Covid-19 and inequality seem to be related (Bambra et al., 2020; Holst et al., 2020). In addition, a study in Germany exposed these financial and health risks associated with Covid-19 and that they were distributed unequally in the early stage of the pandemic (Holst et al., 2020). To elaborate, the measures to reduce the spread of Covid-19 could be seen as a driver for more inequality, such as financial inequality. Yet, the relation between inequality and higher Covid-19 health risks seems the other way around. For example, financial inequality is seen as influence on Covid-19 infection, resulting in more inequality. Concluding, researchers do not see Covid-19 as a great equalizer.

Maestripier (2021) says the Covid-19 pandemic seems to expose existing vulnerabilities, reinforce current inequalities and increase future differences. Moving on, this section will provide some examples of these three developments of inequality for the two channels the Covid-19 pandemic and, financial, inequality seem related. The damage as result of the measures to reduce the spread, and higher health risks for Covid-19.

2.1 The impact of contamination measures

First, the relation between Covid-19 and inequality as result of the social distancing measures will be discussed. The contamination and preventative measures, mostly the social distancing measure, also resulted in negative consequences (Marijn Stok et. al., 2021). It influenced the inequality of health risks, financial situations, academic performances and more.

The already existing vulnerabilities became clearer when a lot of governments undertook measures to reduce the spread of Covid-19. People living in disadvantaged neighborhoods had less capability to uphold the “social distancing measures” (Haase, 2020). Exposing them to a higher health risk of getting Covid-19 and challenging them, financially, more to oblige with the measures. In addition, negative mental health consequences as result of the contamination measures are higher for financial disadvantaged groups (Marijn Stok et. al., 2021). Which is another way the contamination measures expose pre-existing inequalities.

Research in the United States showed that current inequalities get reinforced as a result of the measures. Inequalities in job losses grew across groups with different education levels and demographic characteristics (Montenovo et al., 2020). The study shows that people with higher education levels are facing less unemployment as a result of the Covid-19 pandemic, because work is compatible to be done remotely. Some groups of lower educated workers have actually experienced a lower job security as a result of the pandemic. Yet, some lower educated people have actually also experienced a larger job security, given their jobs in essential industries (Montenovo et al., 2020). Besides job losses and job security, current inequalities also get reinforced through other influences. For example, distancing measures like school closures impacted socioeconomically disadvantaged children more negatively. The academic performances of socioeconomically disadvantaged children are likely to be even more negatively affected by the measures than those of advantaged children (Marijn Stok et. al., 2021). Because, overall, disadvantaged children have less technical and academic support in their "study at home" environment.

To conclude, it is most likely that these developments of the reinforcing of inequalities will result in long-term consequences which will increasing inequality (Haase, 2020). Job losses now could result in financial damage on the long-term and growing differences in academic achievement will most likely lead to more differences in income and jobs.

2.2 Inequalities in Covid-19 infections

Second, Covid-19 and inequalities seem related due to higher health risks for relatively vulnerable people. As discussed, pandemics have caused unequally high rates of contamination and mortality for vulnerable communities in the past (Bambra et al., 2020). Evidence shows that these inequalities are once again present in the Covid-19 pandemic (Abedi et al., 2020).

Before the pandemic started, people living in poverty had significantly more health problems (Bambara et al., 2020). These groups had higher rates of diabetes, asthma, heart diseases, cancer, among other diseases. These unequal rates are a result of unequal exposure to determinants of health, like living conditions, access to essential goods and access to healthcare. The Covid-19 pandemic exposed these inequalities even more, since a lot of these diseases are indicators that increase the severity and mortality of Covid-19. Thus, these unequal distributed health issues lead to unequal distributed Covid-19 health risks.

In addition, current inequalities got reinforced. Lower-paid jobs where much more likely to be designated as key workers (Bambara et. Al., 2020). For example, people working in public transportation or cleaning services. Only the key workers were required to go to work, which meant a higher exposure risk for Covid-19. Earlier describe research also shows that lower educated people got more exposure to health risks due to physical contact at work (Montenovo et al.,2020)

Bambara et al. (2020) also argue that long term consequences of the Covid-19 crisis will be through political and economic pathways. Sudden economic shocks lead to increasing health risks. Resulting in unequal distributed rates of suicide, mental health problems, addiction and morbidity. The pandemic has impacted economies over the world heavily and the "recovery" process could thus influence health risks.

In the investigation of the existing literature, until now, inequality has remained a fairly broad concept. For this thesis research, the average income per area will be used to approach economic inequality. The relation between income and Covid-19 is interesting, because in the past differences in income have proven to lead to inequality in health risks overall (Pickett, 2015).

In addition, the mechanisms behind the infection ratio are analyzed. Average income will be related to multiple Covid-19 ratios of municipalities and areas; the infection ratio, the ratio tests taken and the probability of infection of Covid-19. Borjas (2020) studied these mechanisms with data of New York City in April 2020. This paper will contribute to this existing research by providing more insight on the relations in the Netherlands and within the GGD Rotterdam-Rijnmond region. In addition, understanding how differences in area characteristics relate to the relative number of contaminations, the relative number of tests and the probability of a Covid-19 infection will help contain Covid-19 in a more efficient way.

3. Empirical strategy and data

3.1 Data

3.1.1 Municipality analysis

The relation between average income and Covid-19 infections will be investigated on two levels: municipalities of the Netherlands and the areas of the GGD Rotterdam-Rijnmond region. To answer the research question, data is needed. For the municipality regression data of the National Institute for Public. Health and the Environment (RIVM) and Statistics Netherlands (CBS) is used (National Institute for Public Health, n.d.; Statistics Netherlands, n.d.). The Netherlands counts 355 municipalities in 2020. This research regards 338 municipalities, since there are a few municipalities with missing variables due to changes in the municipality mapping in the past two years.

The data set of RIVM provides the cumulative amount positive Covid-19 tests reported from June 1th 2020 until March 30th 2021, this variable is called P_i. The choice for the use of this specific period is explained at the end of this section. P_i entails the amount of reported COVID-19 patients to the GGD for a given municipality *i*. Covid-19 tests are administered by the GGD, private parties or commercial testers. Private parties administer tests and those results are not always reported to the GGD or the report does not count because the test does not meet the required standards. Commercial testers, for example doctors and laboratories, are obliged to report proven COVID-19 infections to the GGD. In addition, not every possible infected person gets tested. Therefore, the actual amount of infections is higher than reported. Although the figures do not provide a complete picture, they are seen as a good approximation.

The other variables used to analyze the relation between average income and the control variables stem from the CBS. For this research, income data is used as the average yearly personal income per income recipient. In this way, individuals that do not receive any income, such as children, are not considered when calculating the average income. Income can originate from labor, profits from owned companies or social benefits (e.g. unemployment benefits). The variable is called *AverageIncome* and originates from the report *Kerncijfers wijken en buurten 2018* from Statistics Netherlands (n.d.). *Women* is received from CBS as the number of women in a municipality and divided by the total number of inhabitants. *AgeBelow15* is the percentage of inhabitants that are below the age of 15. *AgeAbove65* is the percentage of inhabitants that are above the age of 65. Density is the number of inhabitants per km². *WesternMigrationBackground* and *NonWesternMigrationBackground* are also a percentage. *WesternMigrationBackground* is the total number of residents with a migration background whose origin is one of the countries in the continents of Europe (excluding Turkey), North America, Oceania, Indonesia or Japan, divided by the total number of inhabitants. *NonWesternMigrationBackground* is the total number of residents with a migration background whose origin is one of the countries in the continents of Africa, Latin America and Asia (excluding Indonesia and Japan) or Turkey. Lastly, *IncomeSocialMinimum* is the percentage households with a disposable income up to 110,0% of the social minimum in the year 2019. All above control variables are measured on or before the 1st of January 2020. (Statistics Netherlands, n.d.)

For the municipality analysis the GGD region is used as variable. The variable is also retrieved from CBS and shows to which GGD region a municipality belongs (Statistics Netherlands, n.d.). Municipalities are responsible for managing a total of 25 GGD regions. Several municipalities jointly manage one GGD. All the 25 GGD's together cover the Netherlands. Besides, promoting the health, the GGD's identify and prevent health risks for all its residents.

3.1.2 GGD Rotterdam-Rijnmond analysis

For the regression equations (3), (4) and (5) data of GGD Rotterdam-Rijnmond and Statistics Netherlands (CBS) is used (GGD Rotterdam-Rijnmond, n.d.; Statistics Netherlands, n.d.). The GGD Rotterdam-Rijnmond consists of 15 municipalities, among which Rotterdam. Rotterdam consists of 15 districts, which together contain 93 neighborhoods. Municipality Nissewaard has 85,219 residents, the second highest number of inhabitants after the municipality of Rotterdam. Prins Alexander district has 95,926 residents. If you compare the number of residents in municipality and districts, they do not differ significantly. Therefore, the districts in Rotterdam and municipalities of GGD region Rotterdam-Rijnmond should be highly comparable. Combining both Rotterdam neighborhoods and Rijnmond municipalities creates a total of 29 observations. The research regards 27 areas, since municipality Lansingerland and Rotterdam are excluded.

For the GGD Rotterdam-Rijnmond analysis, data for the Covid-19 rates is retrieved from the GGD Rotterdam-Rijnmond (GGD Rotterdam-Rijnmond, n.d.). This is in contrary to the municipality analysis, where RIVM data is used (National Institute for Public Health, n.d.). P_i is the cumulative number of positive tests with a result that are taken by the GGD's for a given area in the period 1 June 2020 until March 23th 2021. This P_i, in contrast to the municipality analysis, does not included the reported positive tests of commercial parties. Using a P_i where commercial testers are excluded is necessary for conducting an analysis of the mechanisms behind the relative infection rate. Commercial parties are only obliged to report positive tests results to the GGD and they do not report negative or erroneous tests results. For the GGD Rotterdam-Rijnmond analysis the number of total tests T_i, will be used to analyze the mechanisms. T_i is the total number of tests of the residence living in a given area. Since, this information is only available for the test taken by the GGD Rotterdam-Rijnmond P_i and T_i are the positive tests and tests with a result from the GGD Rotterdam-Rijnmond only, excluding the tests taken by commercial and private parties.

The data sources and descriptions for the independent variables are all the same as for the municipality analysis, except from the GGD region – which is not included. Accordingly, all the data of the dependent variables are retrieved from CBS (CBS, n.d.). An overview of all data sources and descriptions for this research can be found in Appendix A, Table 2A.

The RIVM and GGD-Rotterdam-Rijnmond data from June 1th onwards will be used for both the analysis, since from that moment everyone was able to test for free at GGD locations in the Netherlands (Rijksoverheid, n.d.). Area and municipality characteristics are therefore not likely to have affected the allocation of testing resources. Unlike earlier research in neighborhoods in New York

City, where the tests occurred very sporadically (Borjas, 2020). In the beginning of the pandemic testing also occurred very sporadically in the Netherlands, that is why test from that time are not included in this research. Furthermore, the data used for the independent characteristic is all dated between January 1th 2018-2020. Therefore, the possible concern of reversed causality is overcome.

3.2 Methodology

3.2.1 Municipality analysis

To research to what extent differences in economic status influence Covid-19 contaminations of municipalities, an ordinary least squares regression model will be used. The independent variable of interest is *AverageIncome*, which entails the average income of income receivers in a given municipality (i). The dependent variable of interest is the relative infection rate of the municipality. Where P_i indicates the number of positive tests of Covid-19 in given municipality (i) and N_i is the number of people living that municipality (i). The following OLS regression will be estimated:

(1) $$\frac{P_i}{N_i} = \beta_0 + \beta_1 \times AverageIncome_i + \beta_2 \times Women_i + \beta_3 \times AgeBelow15_i + \beta_4 \times AgeAbove65_i + \beta_5 \times Density_i + \beta_6 \times WesternMigrationBackground_i + \beta_7 \times NonWesternMigrationBackground_i + \beta_8 \times IncomeSocialMinimum_i + \beta_9 \times GGDRegion_i + \varepsilon_i$$

As mentioned, the main variable of interest for this paper is average income. The linear regression equation (1) contains control variables to reduce the problem of omitted variables bias (OVB). If these variables would not be included, they would have ended up in the error term. It is important to include particular variables that could influence the dependent variable and might be correlated with the independent variable of interest. Otherwise, the estimated relation of income and the relative infection rate would be biased. The control variables are chosen since existing literature shows that these variables might have explanatory power for Covid-19 infections (Borjas, 2020; Ehlert, 2020).

The control variable *Women* is added because multiple studies have shown different effects for men and women in relation to the Covid-19 pandemic. However, the difference in effect for the genders is still unclear. On the one hand, health risk for women seems smaller, since research suggest that fewer women are dying from Covid-19 than men (Cai, 2020). On the other hand, the measures taken by governments around the world may disproportionately increase the health risks for women (Gausman, 2020). The second control variable is the *AgeBelow15*. Over all, the rule of thumb is that the younger the child, how smaller the risk of spreading the disease (RIVM, 2020). To add, the measures for children, including test indications and school closures, radically changed during the past year (RIVM, n.d.; Rijksoverheid, n.d.). Moreover, people older than 70 year are defined by the Dutch government as high-risk group for Covid-19 (RIVM, n.d.). As a result, the test advice and lockdown measures are stricter. This can result in different relative infection rates among de elderly. Therefore, the variable *AgeAbove65* is added, as data was not available for the age of 70. The control variable *Density* is added because a contagious disease like Covid-19 is expected to spread more rapidly in dense areas. A study on the impact of population density in India shows a moderate association between the spread of Covid-19 and density (Bhadra, 2020). Other findings suggest that larger metropolitan areas have higher infection rates (Hamidi, 2020). Migration background, western and non-western, is added because research has shown that residents in immigrant neighborhoods were less likely to be tested (Borjas, 2020). In contrast, the likelihood that a test was positive was larger for people with a migration background. As explained further in the study, the relative infection rate may depend on these mechanisms. *IncomeSocialMinimum* is added to the regression to control for the income diversity of a municipality. The AverageIncome is aggregated and therefore does not display the variability in income a Municipality has. Lastly, control variable *GGDRegion* is added because different GGD regions can implement differences in policy regarding contract tracing. Moreover, this variable is added to control for geographic differences.

3.2.2 GGD Rotterdam-Rijnmond analysis

The mechanisms behind the relative infection rate will be analyzed in this paper. Area characteristics will be tested on multiple Covid-19 ratios; the infection rate ($\frac{P_i}{N_i}$), the ratio tests in an area ($\frac{T_i}{N_i}$) and the probability of infection of Covid-19 ($\frac{P_i}{T_i}$). These three analyses are performed on the level of the GGD Rotterdam-Rijnmond areas, consisting 14 municipalities and 14 districts of the city Rotterdam. The fraction of people who tested positive for Covid-19 in an area could be written as denoted below (Borjas, 2020):

(2) $$\frac{P_i}{N_i} = \frac{T_i}{N_i} \times \frac{P_i}{T_i}$$

In equation (2), T_i indicates the number of people that tested for Covid-19 in a given area (i). Once again, P_i indicates the number of positive tests of Covid-19 and N_i represents the number of people living that area (i). The equation shows that the fraction of people who tested positive has two underlying mechanism. First, the rate at which the area was tested (T_i/N_i). Second, the rate at which the people who tested, tested positive for Covid-19 (P_i/T_i). If the first ration for the tested area is large, but the rate at which people who tested, tested positive is small, this could balance out. As a result, the relative infection rate will not show the mechanisms clearly. If a population in of a given area does not regularly test for some reason T_i/N_i will be smaller and P_i/T_i bigger. However, not testing does not mean there is no Covid-19. Therefore, only considering P_i/N_i will not give us all the relative information about the Covid-19 spread. Analyzing the two underlying relations will give us extra information about the Covid-19 infections. The following OLS regressions will be estimated:

(3) $$\frac{P_i}{N_i} = \beta_0 + \beta_1 \times AverageIncome_i + \beta_2 \times Women_i + \beta_3 \times AgeBelow15_i + \beta_4 \times AgeAbove65_i + \beta_5 \times Density_i + \beta_6 \times WesternMigrationBackground_i + \beta_7 \times NonWesternMigrationBackground_i + \beta_8 \times IncomeSocialMinimum_i + \varepsilon_i$$

(4) $$\frac{T_i}{N_i} = \beta_0 + \beta_1 \times AverageIncome_i + \beta_2 \times Women_i + \beta_3 \times AgeBelow15_i + \beta_4 \times AgeAbove65_i + \beta_5 \times Density_i + \beta_6 \times WesternMigrationBackground_i + \beta_7 \times NonWesternMigrationBackground_i + \beta_8 \times IncomeSocialMinimum_i + \varepsilon_i$$

(5) $$\frac{P_i}{T_i} = \beta_0 + \beta_1 \times AverageIncome_i + \beta_2 \times Women_i + \beta_3 \times AgeBelow15_i + \beta_4 \times AgeAbove65_i + \beta_5 \times Density_i + \beta_6 \times WesternMigrationBackground_i + \beta_7 \times NonWesternMigrationBackground_i + \beta_8 \times IncomeSocialMinimum_i + \varepsilon_i$$

The same variables are added to regressions (3)-(5) as regression (1) to reduce the problem of OVB. However, not all variables are exactly the same as the first equation. The differences are discussed in the previous section, 3.1 Data. The variable GGD region is excluded since the three regressions are only performed in one region, GGD Rotterdam-Rijnmond.

3.3 Descriptive statistics

3.3.1 Municipality analysis

In this section, descriptive statistics for the infections rate's and municipality characteristics will be provided. The data set is a combination of RIVM (National Institute for Public Health, n.d.) and CBS (Statistics Netherlands, n.d.) data and contains 364 observations. However, the analysis is conducted with 338 observations for which data was available regarding the total population, positive tests and average income.

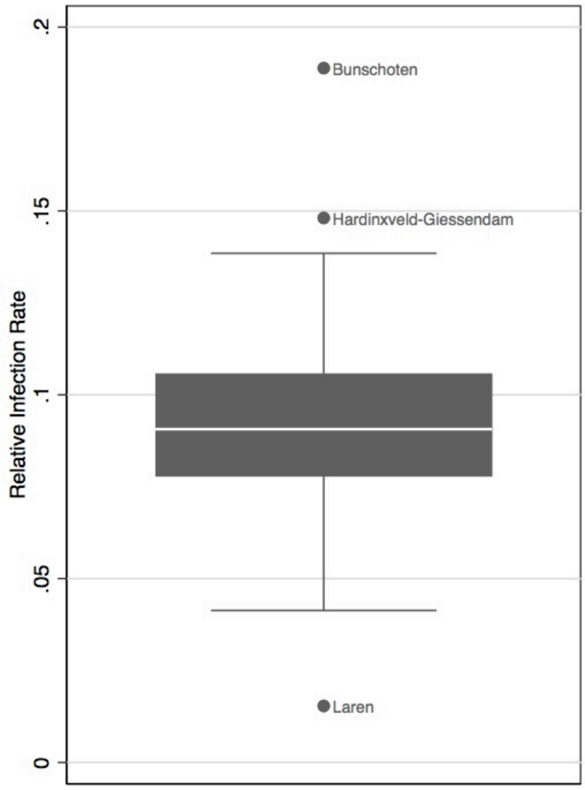
Table 1. Summary statistics municipalities

	(1)	(2)	(3)
Variable	Mean	Min	Max
$\frac{P_i}{N_i}$	0.092	0.015	0.189
	[0.021]		
AverageIncome	32,350.590	24900.00	58,600.00
	[4267.344]		
Women	0.502	0.472	0.527
	[0.008]		
AgeBelow15	0.157	0.102	0.281
	[0.022]		
AgeAbove65	0.215	0.097	0.326
	[0.033]		
Density	903.473	57.000	6620.000
	[1055.032]		
WesternMigrationBackground	0.088	0.016	0.468
	[0.045]		
NonWesternMigrationBackground	0.078	0.014	0.389
	[0.061]		
IncomeSocialMinimum	0.081	0.044	0.188
	[0.026]		
Number of observations	338		

Notes: This table shows the summary statistics of the municipality data. The statistics are constructed with 339 municipalities. Column (1) shows the mean of the variable in the rows. Column (2) the minimum and column (3) of these row variables. The standard deviations are shown between squared brackets.

Table 1 reports summary statistics for municipality characteristics. The infection rate of Covid-19 for municipalities varies between 1,5 and 18,9 percent of the population. Meaning that 1,5 percent to 18,9 percent of a certain municipality population has tested positive on Covid-19 from June 1th 2020 until March 30th 2021. These figures demonstrate the big differences in infection rates across the Netherlands which this thesis research aims to comprehend. Besides, Figure 1 shows the distribution of the Relative Infection Rate. There are three outliers: municipality Laren has a low infection rate and municipalities Hardinxveld-Giessendam and Bunschoten both have a high infection rate.

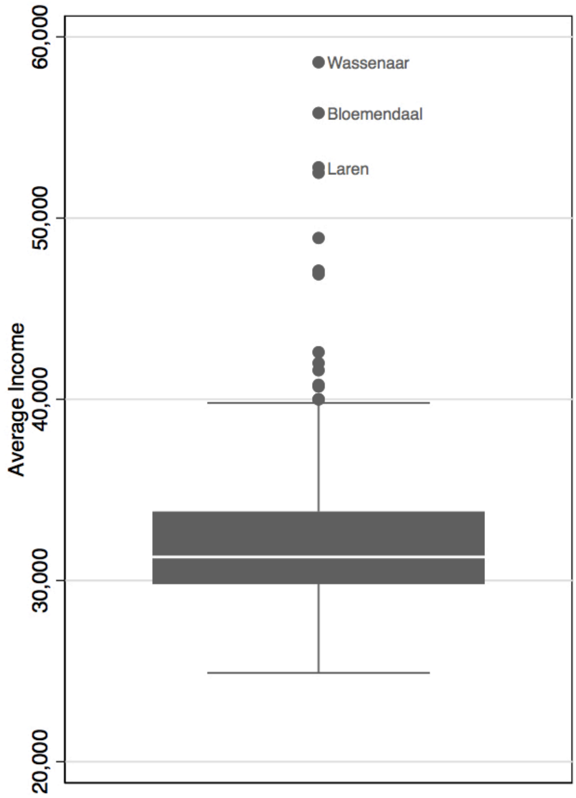
Figure 1. Boxplot Relative Infection Rate Municipalities



Notes: This figure shows the distribution of the Relative Infection Rate, presented in boxplots. The boxplot is constructed with 339 municipalities. The first quartile is equal to 0.078, the median is equal to 0.091 and the third quartile is equal to 0.106. The three outliers are presented with dots and municipality names.

In addition, Table 1 shows a lot of variation between municipality characteristics. The maximum AverageIncome is more than twice as much as the lowest average income of a municipality. Figure 2 shows the distribution of AverageIncome, presented in a boxplot. There are multiple outliers, represented in the dot on the graph. The top three are labeled with municipality names. Moving on, there is around 5 percent point variation in the percentage women living in municipalities (Table 1). The percentage of the population with an age below 15 years varies a lot more than gender, and so does the percentage population at the age of 65 years or older. The population density highly varies between municipalities. The largest population density is found in the municipality of 's Gravenhage with 6620 residents per km², in contrast municipality Grave (Noord-Brabant) has just 57 residents per km². Moreover, on average, municipalities have a relatively larger proportion of people with a western migration background, rather than a non-western migration background. However, the range between the municipality with the least and most percentage residents with a migration background is greater for the percentage residents with a western migration background. This is in contrast to the standard deviation, that is relatively larger for NonWesternMigrationBackground. Lastly, the percentage of households with a disposable income up to 110.0% of the social minimum in 2019 varies between 4.4% and 18.8%.

Figure 2. Boxplot Average Income Municipalities



Notes: This figure shows the distribution of the Average Income for municipalities, presented in boxplots. The boxplot is constructed with 339 municipalities. The first quartile is equal to €29800,-, the median is equal to €31300,-and the third quartile is equal to €33800,-. The outliers are presented with dots and the top three outliers also with municipality names.

Table 2. Summary statistics GGD Region's

	(1)	(2)	(3)
GGD Region	Number of Municipalities	Total Number of Residents	Average Relative Infection Rate
Groningen	6	196.005	0.075
Fryslan	15	602.627	0.063
IJsselland	11	531.342	0.082
Drenthe	12	493.682	0.064
Twente	14	631.064	0.110
Noord- en Oost- Gelderland	22	827.731	0.079
Gelderland-Midden	14	694921	0.083
Gelderland-Zuid	13	510.468	0.099
Utrecht	25	1.298.023	0.094
Hollands Noorden	17	662.748	0.094
Zaanstreeek-Waterland	8	339.182	0.010
Kennemerland	9	549.947	0.085
Amsterdam-Amstelland	6	1.070.575	0.089
Gooi en Vechtstreek	7	308.179	0.075
Haaglanden	9	1.059.041	0.097
Hollands Midden	18	808.860	0.096
Rotterdam-Rijnmond	15	1.323.434	0.104
Zuid-Holland-Zuid	8	328.061	0.121
Zeeland	13	368.586	0.078
West-Brabant	15	652.966	0.088
Hart voor Brabant	24	1.059.041	0.106
Brabant-Zuidoost	21	780.611	0.106
Limburg-Noord	15	520.017	0.110
Zuid-Limburg	15	561.246	0.081
Flevoland	6	423.021	0.083
Number of observations	338		

Notes: This table contains summary statistics for all the 25 GGD regions in the Netherlands. Column (1) displays the number of municipalities in the analysis per GGD region. The municipalities with missing variables are excluded from this overview. The most municipalities where excluded for GGD Groningen. There were 6 municipalities with missing values due to changes in the municipal division and missing values for the islands in the north, resulting in only 6 municipalities n the analysis. For other regions 1 or 2 municipalities where excluded due to the same reasons. Column (2) shows the total amount of residents in a given GGD area. Lastly, column (3) presents the average relative infection rate ($\frac{P_i}{N_i}$) in a given GGD region. Calculated by dividing the number of positive tests by the number of residents in that area.

3.3.2 GGD Rotterdam-Rijnmond analysis

In table 3, descriptive statistics for the GGD areas are provided. GGD (GGD, n.d.) and CBS (Statistics Netherlands, n.d.) data are combined in this data set for the analysis. The data contains 27 observations, since the municipality of Rotterdam and Lansingerland are excluded. A geographical overview of the of the municipalities and districts used in the analysis can be found in Appendix B, Figure 2b.

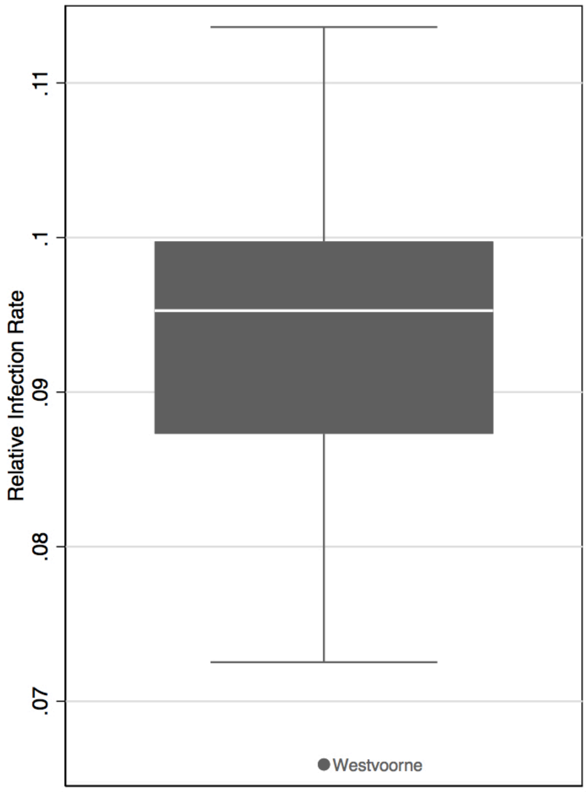
Table 3. Summary statistics GGD Rotterdam-Rijnmond areas

	(1)	(2)	(3)
Variable	Mean	Min	Max
$\frac{P_i}{N_i}$	0.093	0.066	0.114
	[0.012]		
$\frac{T_i}{N_i}$	0.723	0.580	0.864
	[0.081]		
$\frac{P_i}{T_i}$	0.130	0.094	0.179
	[0.021]		
AverageIncome	26159.260	19200.00	33000.00
	[3586.005]		
Women	0.507	0.485	0.524
	[0.009]		
AgeBelow15	0.158	0.090	0.206
	[0.023]		
AgeAbove65	0.189	0.100	0.267
	[0.045]		
Density	4138.481	190.000	14908.000
	[3622.059]		
WesternMigrationBackground	0.109	0.042	0.189
	[0.034]		
NonWesternMigrationBackground	0.235	0.028	0.571
	[0.156]		
IncomeSocialMinimum	0.125	0.055	0.267
	[0.060]		
Number of observations	27		

Notes: This table shows the summary statistics of the GGD Rotterdam-Rijnmond data. The statistics are constructed with 27 areas. Column (1) shows the mean of the variable in the rows. Column (2) the minimum and column (3) of these row variables. The standard deviations are shown between squared brackets.

Table 3 reports that the infection rate of Covid-19 for areas of GGD Rotterdam-Rijnmond vary between 6.6 and 1.4 percent of a given area population. Once again, these rates demonstrate the big differences between infections rates, but in this case even within Rotterdam-Rijnmond. In addition, figure 4 shows the distribution of the Relative Infection Rate for Covid-19. As demonstrated in the boxplot, area Westvoorne is the only outlier. The percentage Covid-19 tests taken by a given population varies between the 58,0 and 86.4 percent and the percentage positive tests of all the tests is on average 13.0 percent. The distribution of these two ratios can be found in appendix B, figures 1B and 2B. Both the percentage Covid-19 tests and percentage positive tests of all tests have no outliers.

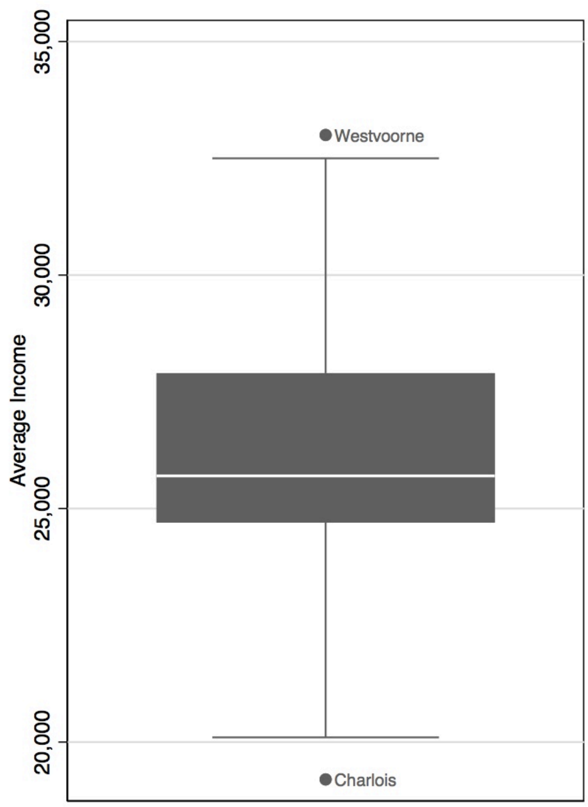
Figure 4. Boxplot Relative Infection Rate GGD area's



Notes: This figure shows the distribution of the Relative Infection Rate for the GGD areas, presented in boxplots. The boxplot is constructed with 27 areas. The first quartile is equal to 0.087, the median is equal to 0.095 and the third quartile is equal to 0.100. The outlier is presented with a dots and area name.

Table 3 also gives a lot of insights about the variation of the characteristics of the areas. The minimum AverageIncome in the GGD Rotterdam-Rijnmond analysis is €19200.00. in contrast the minimum of the municipality analysis is €24900.00 (Table 1). The municipalities included in de GGD Rotterdam-Rijnmond analysis are also included in the municipality analysis. The lower minimum can be explained by the fact that the districts of Rotterdam are not also included in the municipality analysis. This might indicate the relative economic homogeneity within districts – in comparison to municipalities. In fact, it is the district Charlois that has an AverageIncome of €19200.00. In addition, the mean Average Income for the municipality analysis is €32350.59 (Table 1) in comparison to €26159.26 for Rotterdam-Rijnmond (Table 3). Figure 5 shows the distribution of the AverageIncome. There are two outliers: Westvoorne has the highest AverageIncome and Charlois has the lowest AverageIncome. Concluding, area Rotterdam-Rijnmond is not a good representation of the Netherlands in terms of income.

Figure 5. Boxplot Average Income GGD area's



Notes: This figure shows the distribution of the Average Income for the GGD areas, presented in boxplots. The boxplot is constructed with 27 areas. The first quartile is equal to €24700,-, the median is equal to €25700,-and the third quartile is equal to €27900,-. The outliers are presented with dots and area names.

The number of men and women in the GGD Rotterdam-Rijnmond area are almost equal. AgeBelow15 highly differs as some districts might be more child-friendly than others. The GGD Rotterdam-Rijnmond area seems representative for the rest of the country in terms of relative population under 15 years old, as the mean percentage is similar (Table 1; Table 3). AgeAbove65 is on average 18,9 percent of the total population, which is lower than the nation-wide average. The GGD Rotterdam-Rijnmond area is densely populated in comparison to the rest of the country, yet, within the area the population density highly differs. This will be due the fact that Rotterdam has a few very densely populated city districts, while some of the municipalities, such as Goerree-Overflakkee, are not densely populated at all. Furthermore, it is mentionable that Rotterdam has districts with more than 50% residents with a Non-Western migration background. And on average, there are a lot more people with a Non-Western migration compared to people with a Western migration background. This is in exact contrary to the nationwide numbers, showing the cultural diversity of the Rotterdam-Rijnmond area. Lastly, the percentage households with a disposable income up to 110% of the social minimum is on average higher for the GGD Rotterdam-Rijnmond analysis compared to the municipalities

4. Results

4.1 Municipality analysis

The results of regression equation (1) are shown in Table 3. The coefficient of *AverageIncome* on the relative infection rate is -0.002 and statistically significant on a 1% level. Since *AverageIncome* is notated per €1000.00 the interpretation is as followed: if a municipality's *AverageIncome* increases with €1000.00, the relative infection rate ($\frac{P_i}{N_i}$) is expected to decrease with 0.2 percentage points. This is in alignment with the expectations discussed in section 2 Empirical strategy and data. As, presented in section 3.3 Descriptive statistics, the *AverageIncome* from municipalities varies between the €24,900.00 and €58,600.00 (Table 1) and the distribution of *AverageIncome* highly varies (Figure 2). Therefore, an increase of €1000.00 is not uncommon. In addition, the average relative infection rate is 9.2%. So, there is a strongly significant negative correlation of *AverageIncome* on the relative infection rate.

Secondly, Table 3 also presents the coefficients of the added control variables of regression equation (1). An increase of 10.0 percentage points for female residents is estimated to reduce the relative infection rate by 0.7 percentage points. However, an increase of such a magnitude in the number of female residents is very unlikely (Table 1). But most important, the estimated coefficient is not significant. In contrast, the coefficient of the variable of the percentage residents with ages 0-15 is statistically significant on a 1% level. An increase of 10.0 percentage points of the residents with age

0-15 is estimated to lead to an increase of 3.3 percentage points in the relative infection rate. A 10.0 percentage points increase in residents with age 0-15 is likely, and for the average relative infection rate of 9,2%, a 3.3 percentage point increase is a meaningful and a strongly significant difference. Therefore, it can be concluded that the percentage residence with age 0-15 is strongly and significantly correlated with the relative infection rate. The coefficient of the percentage residents with ages 65 and older is not significant. This is also the case for *Density, WesternMigrationBackground, NonWesternMigrationBackground* and *IncomeSocialMinimum*. This is in contrary to the discussed expectations.

Table 3. Regression results for Relative Infection rate, municipality analysis.

	(1)	
Independent variable:	$\frac{P_i}{N_i}$	
AverageIncome	-0.002***	(0.001)
Women	-0.070	(0.163)
AgeBelow15	0.334***	(0.060)
AgeAbove65	-0.03	(0.041)
Density	0.000	(0.001)
WesternMigrationBackground	-0.021	(0.024)
NonWesternMigrationBackground	-0.031	(0.027)
IncomeSocialMinimum	-0.054	(0.058)
GGD Region fixed effect [1]	YES	
Constant	0.178**	(0.071)
Number of observations	338	

Notes: *The results presented in the Table are from regression equation (1). The standard errors are shown between brackets; * p<0.1, ** p<0.05, *** p<0.01. Table 2C in appendix C, shows the result of the same regression excluding the variable IncomeSocialMinimum. The estimated coefficient without IncomeSocialMinimum (Table 2C, Appendix C) is for AverageIncome the exact same as presented in the above Table. However, WesternMigrationBackground and NonWesternMigrationBackground both seem to have statistically significant negative relations with AverageIncome, when IncomeSocialMinimum is not included in the regression equation (Table 2C, Appendix C).*

[1] A table with estimates of effects of all GGD regions can be found in Appendix C, Table 1C.

4.2 GGD Rotterdam-Rijnmond analysis

Table 4 presents the results of the GGD Rotterdam-Rijnmond analysis. The results presented in the table confirm the insights discussed before, that focusing only on the relative infection rate, column (1), falls short of what is really going on with Covid-19 infections. As presented in the first row *AverageIncome* is statistically significant on a 10% level for the relative infection rate. Thus, the *AverageIncome* of a GGD Rotterdam-Rijnmond area is negatively correlated with the relative infection rate of a population. Resulting in a conclusion that Covid-19 infections prevailed more in poor areas than wealthy. However, there is more to it.

Column (2) shows a and statistically significant positive relation, on a 5% level, of *AverageIncome* and the rate at which an area was tested. An increase of €1000.00 in *AverageIncome* is estimated to lead to an increase of 1.0 percentage points in the rate at which an area got tested. Residents in a wealthier area appear to have tested more often than residents of poorer areas. In addition, column (3) reports a statistically significant and negative coefficient for the relation between *AverageIncome* and the chance a person tests positive on Covid-19 on a 1% level. So, if someone lives in an area that has a lower *AverageIncome* of €1000.00, the chance to test positive if tested is estimated to increase with 0.4 percentage points.

Both mechanisms, column (2) and (3), are statistically more significant than the relative infection rate. Therefore, only focusing on the relative infection rate will lead to a weaker conclusion about the correlation, because the results show that the relation between average income and the relative infection rate hides two stronger relationships. If people in different areas would show the same willingness to test, it is highly likely that the correlation between average income and the relative infection rate would be stronger than the estimated 0.2 percent points.

In addition, the investigation of the rate at which an area was tested and the chance a person tested positive on Covid-19 helps to understand the relation between average income and the relative infection rate. An important insight from the results in Table 4 is that the *AverageIncome* and testing ratio are positively correlated, but the *AverageIncome* and the chance to test positive are negatively correlated. This is important to realize, since only evaluating the Covid-19 infection rate does not seem to give a sufficient explanation of the overall spread of Covid-19. When only looking at the infection rate, the conclusion can be made that the relation between *AverageIncome* and Covid-19 infection is simply negative. By extrapolating the equation and having both a positive and negative correlation, for respectively the testing ratio and chance to test positive, policy makers can take grounded actions to better control the Covid-19 spread. For example, through promoting testing in lower income areas or better monitoring the chance to test positive on Covid-19 in a given area to prevent undetected outbreaks. Without these two angles, policy makers could only assume which measurements should be taken.

The estimated relation between the percentage women of an area on the rate at which an area was tested is important to note. The correlation is positive and statistically significant on a 1% level. An increase of 10 percentage points of the percentage women living in an area is estimate to lead to an increase of 34.4 percentage points in the rate at which an area was tested. However, an increase of 10 percentage points is unlikely. But, an increase of 1 percentage point is likely and estimated to lead to an increase of 3.4 percentage points in the relative testing rate. This correlation would not be noticed if the focus was only on the relative infection rate. Apparently, women are more likely to test themselves in the area Rotterdam-Rijnmond. Perhaps this is because women might be more mindful about their symptoms, or because they listen better to the advice of experts or society.

The correlation between the percentage residents with age 65 or older and the rate at which an area was tested, is highly statistically significant on a 1% level. An increase of 10 percentage points in the number of people aged 65+ is estimated to lead to 19.9 percentage points less tested residents. Once again, this correlation would be unnoticed if the focus was solely on the relative infection rate. However, this should be interpreted with the source of the data in mind. The elderly might have been tested more often by commercial parties, because nursing homes where not tested with GGD resources – and the data solely reflects GGD tests. If the tests taken by commercial parties are disproportionally taken by a specific group, it could bias the estimate. This possible limitation is discussed further in the discussion.

Lastly, regression equations (1) from the municipality analysis and equation (3) from the GGD Rotterdam-Rijnmond analysis are comparable since they use the exact same variables, except for the exclusion of the GGD Region variables from equation (3). But, most importantly, they differ in their observations. Regression equation (1), presented in Table 3, uses 338 municipality in the Netherlands as observations and equation (3) 27 areas in GGD Rotterdam-Rijnmond. A lower number of observations implies less power, such that the relations are less precisely estimated. For example, the estimate of *AverageIncome*, column (1), is only statistically significant on a 10% interval. Yet, the municipality estimate in Table 3 is statistically significant on an 1% level. Noticeable, the two analysis have the same estimated coefficient for the relation between *AverageIncome* and the relative infection rate. Both results, Table 3 and 4, show a negative estimate of 0.2 percentage points. In addition, the estimate of *AgeBelow15* changed direction and is not statistically significant anymore. Furthermore, a lot of coefficients that were not statistically significant in Table 3 changed and remained not statistically significant in column (1), table 4.

Table 4. Regression results for the GGD Rotterdam-Rijnmond analysis.

	(1)	(2)	(3)
Independent variable:	$\frac{P_i}{N_i}$	$\frac{T_i}{N_i}$	$\frac{P_i}{T_i}$
AverageIncome	-0.002*	0.010**	-0.004***
	(0.002)	(0.004)	(0.001)
Women	0.347	3.436*	-0.048
	(0.245)	(1.832)	(0.400)
AgeBelow15	-0.074	0.035	-0.072
	(0.107)	(0.709)	(0.159)
AgeAbove65	-0.089	-1.986***	0.222
	(0.107)	(0.380)	(0.142)
Density	0.001	-0.000	0.001
	(0.001)	(0.008)	(0.002)
WesternMigrationBackground	0.051	0.287	0.015
	(0.114)	(0.550)	(0.156)
NonWesternMigrationBackground	0.020	-0.556	0.143
	(0.061)	(0.456)	(0.104)
IncomeSocialMinimum	-0.075	0.727	-0.240
	(0.118)	(0.791)	(0.185)
Constant	-0.023	-0.911	0.209
	(0.114)	(0.751)	(0.170)
Number of observations	27		

Notes: *Column (1) presents the results of regression equation (3), where the relative infection rate is the dependent variable. Column (2) shows the estimate coefficients of regression equation (4), with the rate at which the area was tested is used as the dependent variable. The last column shows the results of regression equation (5) with the rate at which the people who tested, tested positive for Covid-19 as dependent variable. The standard errors are shown between brackets; * p<0.1, ** p,<0.05, *** p<0.01. Table 3C in Appendix C shows the same result excluding the variable IncomeSocialMinimum. The estimate of AverageIncome on the relative infection rate is not significant when the variable is excluded.*

5. Discussion

It is important to interpret the results found in the analysis. First, the magnitude of the relations estimated. The impact of the -0.2 percentage points as result of an increase of €1000.00 in average income and the relative infection rate, can be approached in different ways. At first, a medical approach: an unequal increase in the relative infection rate is most likely leading to an unequal increase in the number of diseased Covid-19 patients. This approach considering, the magnitude of the relation found in this research seems small. The percentage deceased as result of Covid-19 is difficult to estimate. However, RIVM reports a total of 17,775 deaths as result of Covid-19 (RIVM, n.d.), which is only a fraction of all the Covid-19 infections. Therefore, a decrease of 0.2 percentage points if income increases with €1000.00, does probably not lead to a lot more people dying due to Covid-19. Secondly, the subject can be viewed economically and socially. Measured Covid-19 infections lead to quarantine and isolation measures. From this view the impact of a change of 0.2 percentage points is larger. One infection on average leads to quarantine advise for 2 household members, 2 close contacts and 3 to 9 other contacts. (RIVM, 2021). This number of contacts is strongly influenced by the changes in contaminations measures taken by the government at a specific time. The infected person and contacts are all expected to stay home for minimum of 5 and up to 14 days. This can have a significant impact on someone's life. People sometimes cannot work from home, have to spend extra money on home deliveries and people suffer from the lack of social contact. All besides the physical health risk incurred.

In addition, the results coming from the investigated mechanisms should be interpreted. The statistically significant positive relation between average income and the rate at which an area was tested could be due to a broad range of aspects. Relatively poor people could for example test less because it is more difficult for them to find transportation to test locations. The use of a cars is encouraged and using public transportation is discouraged by the government. This can make it harder for poor people to get tested. Another potential explanation of this negative correlation could be that residents with a relatively high income often work from home during the pandemic and can easily get tested during working hours. For key workers, who have on average a lower income, working from home is not feasible and thus, testing during working hours might be difficult. This lower test rate could also possibly explain the negative relation found between average income and the chance a person tests positive for Covid-19. If people do not know they are infected, because they do not test, they continue their daily business while being contagious. This will expose people in lower income areas a lot more to the risks of getting and spreading Covid-19 and can increase the percentage positive tests for relatively poor areas. In addition, the chance a person tests positive for Covid-19 could also be higher in relatively poor areas because of other factors. For example, in section 2, related literature, it is discussed that low-income groups have higher health risks due to more exposure at work or poorer housing situations.

The above considerations will result in the fact that the reported Covid-19 infection rate is less representative for the actual infection rate for lower income areas, compared to higher income areas. The number of actual infections in these lower income areas will presumably be a lot higher than the reported number of infections. This also allows one to say that the risks for inhabitants are not equally divided among areas. Therefore, it is important from an efficiency and equality perspective to let policy makers focus on areas with relatively low incomes.

What can be done, is improving the ability to test in low income areas. For example, free testing possibilities at local markets in those poor areas, but also better information management to inhabitants of these areas. Another option could be placing the testing locations in the center of the areas, rather than large buildings in the outskirts of the areas (e.g. SS Rotterdam, Ahoy or Van Nelle Fabriek). Furthermore, schools are a central hub in areas and could function as a testing location for both parents and children.

If the test rate would be the same among all the areas, Covid-19 infection rates between areas would be better comparable. In this situation, it is highly likely that there is still a significant difference in infection rates between these areas, given that the chance of testing positive for Covid-19 will still differ. Potential explanations for these differences have been discussed. Yet, this research cannot explain all the difference in the chance to test positive for Covid-19 set of by income inequality. This research mostly shows that testing is essential to accurate decision making for policy makers and that unequal testing rates as result of income inequality will lead to unequal health risk and inefficient Covid-19 contamination.

Moving on some limitations of the method and data use will be discussed. Both the municipality and GGD Rotterdam-Rijnmond analysis show significant negative relation between the average income and the relative infection rate. Even the mechanisms behind the infection rate have proven to show significant relations with average income for the GGD Rotterdam-Rijnmond areas. Average income and the testing ratio are positively correlated and average income and the chance to test Covid-19 positive are negatively correlated. However, these results have potential confounding factors that are not completely controlled for by the added control variables. Therefore, the results should be interpreted as correlations and not as causal effects. For example, in the investigation of related literature, section 2, it is briefly mentioned that people living in disadvantaged neighborhoods had less capability to uphold the "social distancing measures" (Haase, 2020). A confounding factor that could be correlated with the relative infection rate and the average income could for example be the type of housing in municipalities and areas. People with less income are for example expected to live in smaller houses, which could play a role in a higher contamination risk. This issue of endogeneity could cause the coefficient of average income on the relative infection rate to be overestimated. In further research a variable like "type of housing" could be added to the regressions to reduce the omitted variable bias. However, it is hard to overcome the omitted variable bias completely. In addition, an experiment with Covid-19 infections and average income is most likely to be unethical and other research methods require time or an ex post approach, which is not very practical in the current evolving Covid-19 crisis. Therefore, the regression method used in this research is an adequate approach to answerer the stated research questions.

The 27 areas used as observations in the ordinary least squares regression of the GGD Rotterdam-Rijnmond analysis is very low. The number of observations used in a regression influences various measures. For example, the average, range and measures of variance. Most importantly a low number of observations reduces the statistical power and validity of an analysis. The relation between average income and the relative infection rate of the GGD Rotterdam-Rijnmond analysis is only statistically significant on a 10% level and the validity is restricted to the specific 27 areas. On its own the GGD Rotterdam-Rijnmond analysis might not seem a significant contributing to the literature. Since the low number of observations comes with implications. However, the municipality analysis has a lot more observations and supports the claims made with the GGD Rotterdam-Rijnmond analysis. Noticeable, the relation between average income and the relative infection rate is exactly the same for both analysis, -0.2 percentage points. The only difference between the two is that the estimate of the municipality analysis is statistically significant on a 1% level, in contrast to a 10% level. However, there was no testing variables available for the municipality analysis to research the underlying mechanism. Therefore, the municipality analysis can't support the GGD Rotterdam-Rijnmond analysis on the results of the mechanisms. But, it is most likely that the same relations for the mechanisms on municipality level will be found if investigated.

An important note to make is that the data used for the regressions is aggregated. First, the independent variables are averages from municipalities and areas. The variation of the variables in those observation highly differs; some municipalities of areas are more divers. The variable of the percentage of households with a disposable income up to 110.0% of the social minimum is added to control for this diversity problem. However, this variable is only an approach for the lower half of the income distribution and not the total diversity in a municipality or are. Moreover, the three Covid-19 ratios used in the analysis are all based on cumulative numbers from June 2020 up to May 2021. Therefore, the results found in this paper are not very specific and are for example not suited to for managing an outbreaks day to day. The results can help by overall policy decisions and implementing measures to reduce the spread of Covid-19.

The allocation and location of testing resources could influence the number of positive tests and tests taken by the GGD Rotterdam-Rijnmond. First of all, the location of testing. A low accessibility or large distance to the testing locations can reduce the testing rate of specific areas and municipalities. As shown in this paper, the testing rate can influence the relative infection rate. Moreover, the allocation of testing can also be problematic. For example, a lot of nursing homes get testing resources arrangements independently from the GGD. These parties are only obliging to communicate positive tests with the GGD. Therefore, all the negative Covid-19 tests taken of this population group are not reported. This can lead to an overestimation of the relations. The results in Table 4 report that residents with age 65 or older have a statistically significant lower relative testing rate. This result can be biased by the testing allocation. However, it is really hard to overcome the problem of allocation and location of testing resources. The number of tests taken by all testing providers are not available and the accessibility to testing locations is subjective, which makes it hard to control for these problems.

6. Conclusion

The Covid-19 crisis has caused countries to implement measures to minimize the spread of Covid-19. Many policy makers and researchers try to grasp the health, economic and social effects of Covid-19. Existing research provides evidence that Covid-19 might lead to increasing inequality in society. Which is in alignment with literature about the relation between health and income. However, research of past pandemics display that past pandemics redistributed towards the poor or exterminated the poor. To elaborate excising literature and provide tangible insight about the spread, this paper studies the effect of average income on the relative Covid-19 infection rate. The hypothesis is that, if municipalities or areas have less average income the relative infection rate will be higher.

To investigate, municipalities and areas in GGD Rotterdam-Rijnmond are analyzed. The results of the linear regressions show that there is a statistically significant negative relation between average income and the relative infection rate. Therefore, a decrease in the average income of an area is expected to lead to an increase in the Covid-19 infection rate. Secondly, the GGD Rotterdam-Rijnmond analysis shows a significant positive relation between average income and the rate at which an area was tested and a significant negative relation between average income and the chance a person tests positive on Covid-19. Therefore, residents in poor areas appear to have tested for Covid-19 at a lower rate and have higher chances of testing positive if tested than residents of richer areas. The results provide information on correlations between the variables, but not on causal effects.

The results have important policy implications. For example, it can help politicians and policy makers explain differences in relative infection rates. The research shows that differences in the relative infection rate can come from differences in income. In addition, the knowledge that areas with less income also have a lower testing rate can help enhance the effectiveness of "Covid-19 testing campaigns" by targeting areas differently. Moreover, the given insights of this research can help to make policy makers aware of the underlying mechanism of the relative infection rates. Policy makers should highly focus on lower income areas, to improve both testing rates and control measurement, to better get Covid-19 under control. This would be beneficial from an equality and efficiency point of view. Further research could focus on those mechanisms to investigate if the results found in this research also can be found for all municipalities or other specific regions, like neighborhoods.

References

Abedi, V., Olulana, O., Avula, V., Chaudhary, D., Khan, A., Shahjouai, S., Li, J., & Zand R. (2020). Racial, Economic, and Health Inequality and COVID-19 Infection in the United States. *Journal of Racial and Ethnic Health Disparities.* https://link.springer.com/article/10.1007/s40615-020-00833-4

Alfani, G. (2020). Epidemics, inequality and poverty in preindustrial and early industrial times. *Stone Centre on Socio-Economic Inequality.*

Bambra, C., Riordan, R., Ford, J., & Matthews, F. (2020). The COVID-19 pandemic and health inequalities. *Journal of Epidemiology and Community Health.* https://jech.bmj.com/content/jech/74/11/964.full.pdf.

Bhadra, A., Mukherjee, A., & Sarkar, K. (2021). Impact of population density on Covid-19 infected and mortality rate in India. Modeling Earth Systems and Environment, 7(1), 623-629.

Borjas, G. (2020). Demographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods. *National Bureau of Economic Research.* https://www.nber.org/system/files/working_papers/w26952/w26952.pdf.

Bowleg, L. (2020). We're Not All in This Together: On COVID-19, Intersectionality and Structural Inequality. *American Journal of Public Health.* https://ajph.aphapublications.org/doi/pdfplus/10.2105/AJPH.2020.305766

Cai, H. (2020). Sex difference and smoking predisposition in patients with COVID-19. The Lancet Respiratory Medicine, 8(4), e20.

Ehlert, A., (2020). The socio-economic determinants of COVID-19: A spatial analysis of German country level data. *MedRxiv. https://www.medrxiv.org/content/medrxiv/early/2020/07/07/2020.06.25.20140459.full.pdf*

Gausman, J., & Langer, A. (2020). Sex and gender disparities in the COVID-19 pandemic. *Journal of women's health,* 29(4), 465-466.

GGD Rotterdam-Rijnmond (n.d.). *Gezondheidinkaart*. https://gezondheidinkaart.nl/jive?workspace_guid=6df2221e-2083-41dd-80a0-6b01f304330b.

Göran, D., & Whitehead, M. (1991). Policies and strategies to promote social equity in health. *http://repositori.uji.es/xmlui/bitstream/handle/10234/187797/GoeranD_Policies_and_strategies_to_promote_social_equity_in_health.pdf?sequence=1*

Haase, A. (2020). Covid-19 as a social crisis and justice challenge for cities. Frontiers in Sociology, 5. Chicago

Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. Journal of the American Planning Association, 86(4), 495-509.

Holst, H., Fessler, A., & Niehoff, S. (2020). Covid-19, social class and work experience in Germany: inequalities in work-related health and economic risks. *European Societies.* https://doi.org/10.1080/14616696.2020.1828979.

Instiruut Fysieke Veiligheid (n.d.). Gezondheidszorg 25 GGD Regio's. Retrieved https://www.ifv.nl/kennisplein/Paginas/Kaart-GGD-regios-2018.aspx

Maestripieri, L. (2021). The Covid-19 Pandemics: why Intersectionality Matters. *Frontiers in Sociology.* https://www.frontiersin.org/articles/10.3389/fsoc.2021.642662/full.

Montenovo, L., Jiang, X., Lozano Rojas, F., Schmutte, I., Simon, K., Weinberg, B., & Wing, C. (2020). Determinants of Disparities in Covid-19 Job Losses. *National bureau of Economic Research*. https://www.nber.org/system/files/working_papers/w27132/w27132.pdf.

National Institute for Public Health and the Environment (n.d.). *Children, school and Covid-19.* https://www.rivm.nl/coronavirus-covid-19/kinderen

National Institute for Public Health and the Environment (n.d.). *Covid-19 cumulatieve aantallen per gemeente https://data.rivm.nl/geonetwork/srv/dut/catalog.search#/metadata/1c0fcd57-1102-4620-9cfa-441e93ea5604.*

National Institute for Public Health and the Environment (June 2021). *BCO bij een toenemende vaccinatiegraad*. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwj38vyG3uzxAhUSLewKHb-DAe0QFjABegQICxAD&url=https%3A%2F%2Fwww.rijksoverheid.nl%2Fbinaries%2Frijksoverheid%2Fdocumenten%2Fpublicaties%2F2021%2F06%2F18%2Fbco-bij-een-toenemende-vaccinatiegraad%2Fbco-bij-een-toenemende-vaccinatiegraad.pdf&usg=AOvVaw0KepNz3YKr-5IMDV7dfsLD

Statistics Netherlands (n.d.). *StatLine* *https://opendata.cbs.nl/#/CBS/nl/dataset/84583NED/table?ts=1623418964914*

Rijksoverheid (n.d.). *Coronavirus tijdlijn.* *https://www.rijksoverheid.nl/onderwerpen/coronavirus-tijdlijn/juni-2020-versoepeling-coronamaatregelen-en-testen-voor-iedereen*

Pickett, K. E., & Wilkinson, R. G. (2015). Income inequality and health: a causal review. Social science & medicine, 128, 316-326.

Appendix A

Table 1A: *Municipality analysis variable*

Variable	Source	Explanation
Municipalities	CBS	The list of municipalities in the Netherlands, accumulating to a total of 355 in 2021.
GGD Region	CBS	Each of the 355 municipalities belong to a GGD Region, with a total number of 25 regions.
Total positive tests	RIVM	Total cumulative number of positive tests retrieved from RIVM. The positive test are cumulative from the period June 1st 2020 until May 30th 2021.
Total population	GGD	Number of residents living in a given area in 2020
Density	GGD / CBS	Number of residents per km² on January 1th 2020.
Percentage age 65+	CBS	Number of residents with an age 65 or higher living in a given area on January 1th 2020, divided by the total population of the given area.
Percentage of Woman	CBS	Number of women living in a given area on January 1th 2020, divided by the total population of the given area.
Percentage age 0-15	CBS	Number of residentss with the age 0-15 living in a given area on January 1th 2020.
Amount MigrationBackground Western	CBS	Number of residentss with a migration background whose origin is one of the countries in the continents of Europe (excluding Turkey), North America and Oceania or Indonesia or Japan living in the area on January 1th 2020.
Amount MigrationBackground Non-Western	CBS	Number of residents with a migration background whose origin is one of the countries in the continents of Africa, Latin America and Asia (excluding Indonesia and

		Japan) or Turkey living in the area on January 1th 2020.
Average income per income recipient	CBS	The average income per income recipient is calculated as the average income of a neighborhood or municipality, per person that is obtaining income.
110% Social minimum income	CBS	The percentage households with a disposable income up to 110,0% of the social minimum in the year 2019 for a given municipality

Table 2A: GGD Rotterdam-Rijnmond analysis variables

Variable	Source	Explanation
Area's Rotterdam-Rijnmond	CBS, GGD	The Area's of GGD Rotterdam-Rijnmond used for this research consist of 14 municipalities of which one is the municipality is Rotterdam – split up in 14 neighborhoods in this research.
Total tests (with result)	GGD, Coroon IT	Total number of tests administered by the GGD with a result from week 23 in 2020 to week 20 in 2021: June 1st 2020 until May 23th 2021.
Total positive tests	GGD, Coroon IT	Total number of tests administered by the GGD with a positive result from week 23 in 2020 to week 20 in 2021: June 1st 2020 until May 23th 2021.
Total population	GGD	Number of residents living in a given area in 2020
Density	GGD / CBS	Number of residents per km² on January 1th 2020.
Percentage age 65+	CBS	Number of residents with an age 65 or higher living in a given area on January 1th 2020, divided by the total population of the given area.
Percentage of Woman	CBS	Number of woman living in a given area on January 1th 2020,

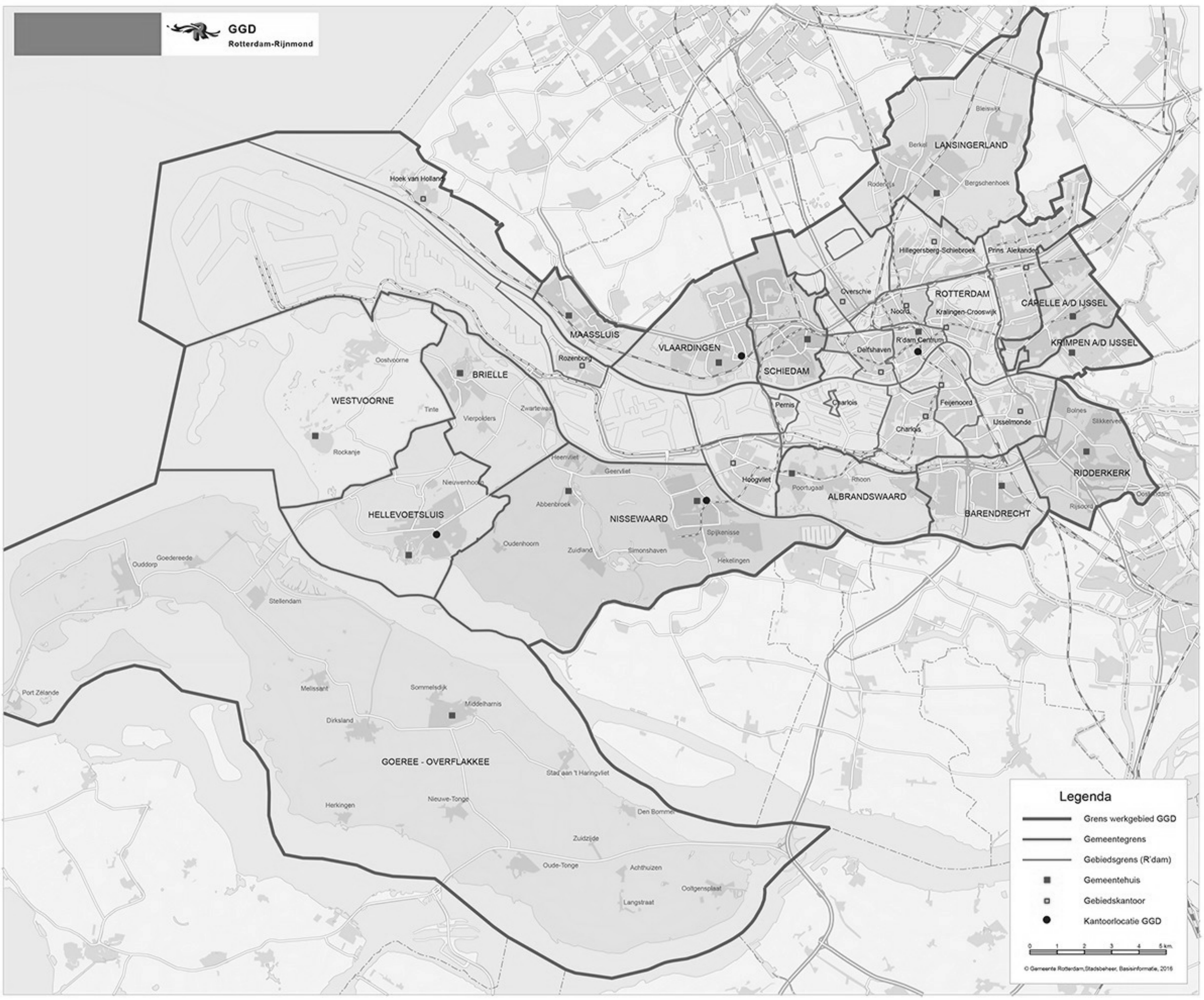
		divided by the total population of the given area.
Percentage age 0-15	CBS	Number of residentss with the age 0-15 living in a given area on January 1th 2020.
Amount MigrationBackground Western	CBS	Number of residentss with a migration background whose origin is one of the countries in the continents of Europe (excluding Turkey), North America and Oceania or Indonesia or Japan living in the area on January 1th 2020.
Amount MigrationBackground Non-Western	CBS	Number of residentss with a migration background whose origin is one of the countries in the continents of Africa, Latin America and Asia (excluding Indonesia and Japan) or Turkey living in the area on January 1th 2020.
Average income per income recipient	CBS	The average income per income recipient is calculated as the average income of a neighborhood or municipality, per person that is obtaining income.
110% Social minimum income	CBS	The percentage households with a disposable income up to 110,0% of the social minimum in the year 2019 for a given municipality

Appendix B

Figure 1B. *Geographic overview of all the GGD Regions of the Netherlands*

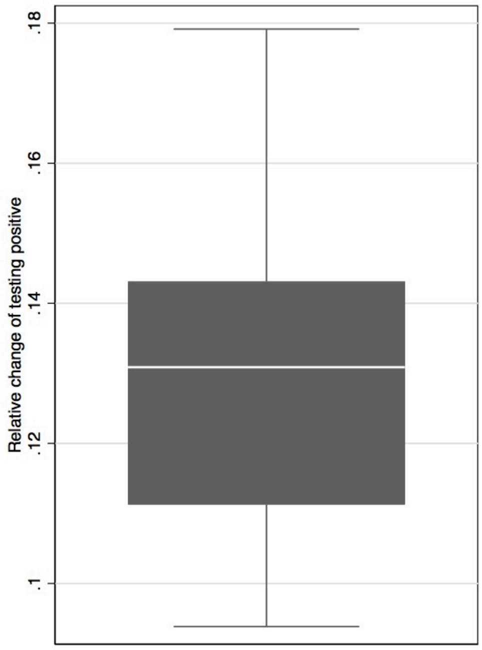
Notes: This geographic figure is retrieved from Institute Fiscal safety (IFV, n.d.). The map shows the division of 25 GGD Regions in the Netherlands.

Figure 2B. *Geographic overview of all the Area's in the GGD Rotterdam-Rijnmond analysis*



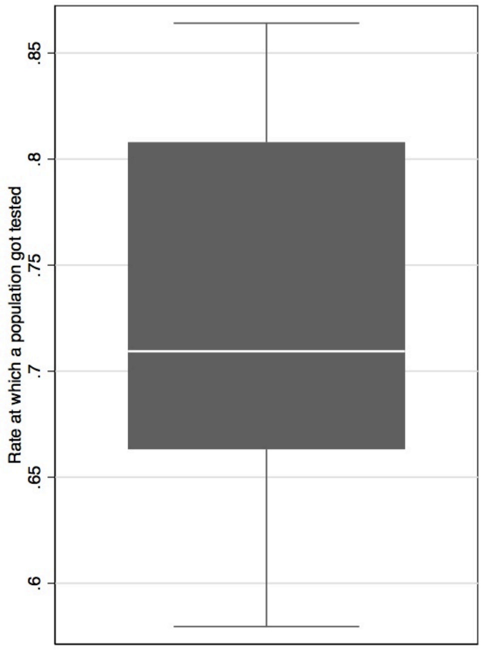
Notes: This geographic overview is retrieved from GGD Rotterdam-Rijnmond (Gezondheidsinkaart, n.d.). The map shows the division areas in the region GGD Rotterdam-Rijnmond.

Figure 1B. Boxplot change of testing positive, GGD areas



Notes: This figure shows the distribution of the change of testing positive for Covid-19, if tested, for the GGD areas, presented in a boxplot. The boxplot is constructed with 27 areas. The first quartile is equal to 0.111, the median is equal to 0.131 and the third quartile is equal to 0.143. There are no outliers.

Figure 2B. Boxplot rate at which GGD areas where tested



Notes: This figure shows the distribution of the rate at which GGD areas got tested on Covi-19, presented in boxplots. The boxplot is constructed with 27 areas. The first quartile is equal to 0.663, the median is equal to 0.709 and the third quartile is equal to 0.808. There are no outliers.

Appendix C

Table 1C. Regression results for Relative Infection rate with fixed effects, municipality analysis.

	(1)	
Independent variable:	$\frac{P_i}{N_i}$	
AverageIncome	-0.002***	(0.001)
Women	-0.070	(0.163)
AgeBelow15	0.334***	(0.060)
AgeAbove65	-0.03	(0.041)
Density	0.000	(0.001)
WesternMigrationBackground	-0.021	(0.024)
NonWesternMigrationBackground	-0.031	(0.027)
IncomeSocialMinimum	-0.054	(0.058)
GGD Groningen	-0.039***	(0.007)
GGD Fryslan	-0.054***	(0.004)
GGD Ijsselland	-0.041***	(0.006)
GGD Drenthe	-0.047***	(0.006)
GGD Twente	-0.006	(0.005)
GGD Noord- en Oost- Gelderland	-0.033***	(0.005)
GGD Gelderland-Midden	-0.027***	(0.005)
GGD Gelderland-Zuid	-0.011*	(0.006)
GGD Utrecht	-0.012**	(0.005)
GGD Hollands Noorden	-0.014***	(0.006)
GGD Zaanstreek-Waterland	-0.004	(0.006)
GGD Kennemerland	-0.010*	(0.005)
GGD Amserdam-Amstelland	-0.003	(0.006)
GGD Gooi en Vechtstreek	-0.013	(0.009)
GGD Haaglanden	-0.001	(0.006)
GGD Hollands-Midden	-0.008*	(0.004)
GGD Rotterdam-Rijnmond	-	
GGD Zuid-Holland-Zuid	0.007	(0.005)
GGD Zeeland	-0.031***	(0.006)
GGD West-Brabant	-0.016***	(0.005)
GGD Hart voor Brabant	-0.002	(0.005)
GGD Brabant-Zuidoost	-0.000	(0.005)

GGD Limburg-Noord	0.005	(0.006)
GGD Zuid-Limburg	-0.016***	(0.006)
GGD Flevoland	-0.042***	(0.007)
Constant	0.178**	(0.071)
Number of observations	338	

Notes: *the standard errors are shown between brackets; * p<0.1, ** p,<0.05, *** p<0.01.*

Table 2C. Regression results Relative Infection rate, municipality analysis, excluding IncomeSocialMinimum.

	(1)	
Independent variable:	$\frac{P_i}{N_i}$	
AverageIncome	-0.002***	(0.000)
Women	-0.094	(0.154)
AgeBelow15	0.336***	(0.062)
AgeAbove65	-0.035	(0.042)
Density	0.000	(0.001)
WesternMigrationBackground	-0.032*	(0.020)
NonWesternMigrationBackground	-0.045**	(0.025)
GGD Region fixed effect	YES	
Constant	0.184**	(0.069)
Number of observations	338	

Notes: *The results presented in the Table are from regression equation (1), exluding the variable IncomeSocialMinimum.. The standard errors are shown between brackets; * p<0.1, ** p,<0.05, *** p<0.01.*

Table 3C. Regression results GGD Rotterdam-Rijnmond analysis, excluding IncomeSocialMinimum.

	(1)	(2)	(3)
Independent variable:	$\frac{P_i}{N_i}$	$\frac{T_i}{N_i}$	$\frac{P_i}{T_i}$
AverageIncome	-0.001	0.010**	-0.003**
	(0.001)	(0.003)	(0.001)
Women	0.230	3.925*	-0.209

	(0.215)	(1.451)	(0.339)
AgeBelow15	-0.074	0.040	-0.073
	(0.109)	(0.732)	(0.176)
AgeAbove65	-0.093	-1.948***	0.209
	(0.112)	(0.380)	(0.162)
Density	0.001	-0.000	0.001
	(0.001)	(0.008)	(0.002)
WesternMigrationBackground	0.053	-0.310	-0.016
	(0.042)	(0.286)	(0.167)
NonWesternMigrationBackground	0.002	-0.310	0.106
	(0.104)	(0.287)	(0.075)
Constant	-0.002	-1.128	0.281
	(0.104)	(0.573)	(0.151)
Number of observations	27		

Notes: *Column (1) presents the results of regression equation (3), where the relative infection rate is the dependent variable. Column (2) shows the estimate coefficients of regression equation (4), with the rate at which the area was tested is used as the dependent variable. The last column shows the results of regression equation (5) with the rate at which the people who tested, tested positive for Covid-19 as dependent variable. All three regression equations presented in the graph are exluding the variable IncomeSocialMinimum. The standard errors are shown between brackets; * p<0.1, ** p,<0.05, *** p<0.01.*