Exploring socio-demographic factors on vaccination uptake in the Netherlands: A Geographically Weighted Regression model

Mathijs van der Kroft, Sebastiaan van Rijk, Thomas Mernagh

Abstract

This project investigates the socio-demographic factors behind possible SARS-CoV-2 vaccine hesitancy. A focus on a country like The Netherlands, with considerable diversity, allows for extensive review of varied socio-demographic factors. Investigation was done at both neighbourhood and municipality level. A spatial lag and a fixed kernel and Geographically Weighted Regressions (GWR) model was applied to the data. The spatial lag model provided a R2 value of 0.59 at neighbourhood level and an AIC of 5752.7. The spatial lag model provided a R2 value of 0.59 at municipalities level and an AIC of 579.6. A Geographically Weighted Regression model was fitted at both neighbourhood and municipality level to estimate model coefficients and global R2 values. Municipality showed a global R2 of 0.66 and neighbourhood showed a value at 0.56. Results demonstrated the ability to use open-source data to find relations in demographic factors lending to vaccine hesitancy in the Netherlands.

Introduction

It is important to ensure that all groups have access to the SARS- CoV- 2 vaccine, especially disadvantaged groups who are more susceptible to infection from SARS-CoV-2 and at greater risk of severe morbidity and mortality. Where vaccines are available, it remains to be seen whether those in greatest need and most affected by the pandemic will be willing and able to access the vaccine. A successful vaccination campaign requires three things: vaccine supplies (and appropriate storage and distribution), people to implement them (vaccinators), and willingness of people to be vaccinated (1). In the Netherlands, at the time of writing, there are enough vaccine supplies and enough vaccinators to vaccinate everyone. Meaning the only determining factor in the uptake of the vaccine was the willingness to be vaccinated. This is the interest of this paper; can we find socio-demographic factors explaining the vaccination coverage per region and thus explain willingness to be vaccinated in the region.

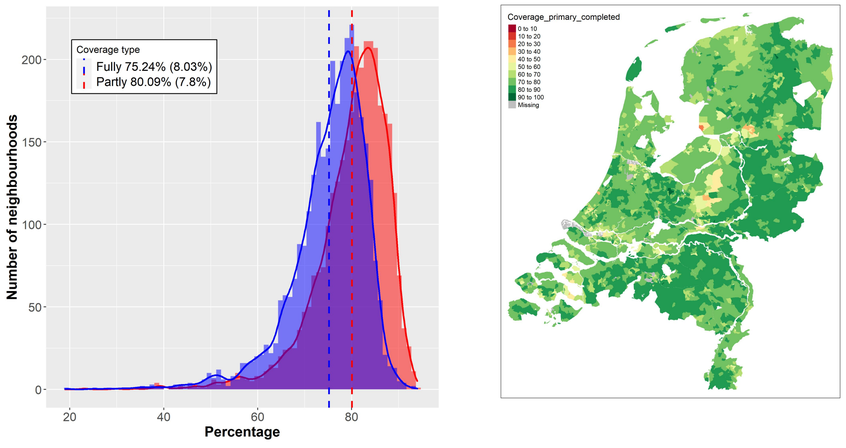


Fig. 1. Vaccination coverage against SARS-CoV-2 in the Netherlands on march 14th 2022

Methodology

Data wrangling

Merge data sources
- Vaccine data
- Polygons (spatial data)
- CBS statistics

Clean data
- Remove (hidden) NA's
- Remove 0 inhabitants
- Z-Normalize

Linear model
Select predictors on:
- High significance
- High coëfficients
- High R2
- Low multicollinearity

Spatial lag model
- Best weights matrix
  - IDW
  - contingency
- Refine predictor selection on
  - Moran's I

GWR model
Select kernel shape and fixed/adaptive bandwidth based on:
- low AIC
- High quasi-global R2

Refine predictor selection based on:
- Low variable coefficient variance

Results

Neighbourhoods

Spatial lag R2 = 0.59, AIC = 5753

Variable	Estimate	SD	p-value
Intercept	+0.005	0.011	0.68
15 to 25 years	**-0.182**	**0.016**	**2.2e-16**
65 years or older	**+0.119**	**0.016**	**9.8e-14**
Western migration	**-0.152**	**0.015**	**2.2e-16**
Non-western migration	**-0.305**	**0.016**	**2.2e-16**
Relative births	**-0.113**	**0.019**	**1.4e-9**
Total households	**+0.070**	**0.013**	**3.2e-8**
Low education	**-0.086**	**0.018**	**1.2e-6**
High education	**+0.260**	**0.018**	**2.2e-16**

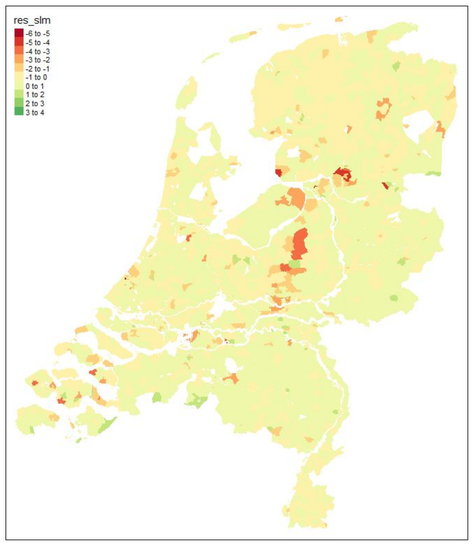
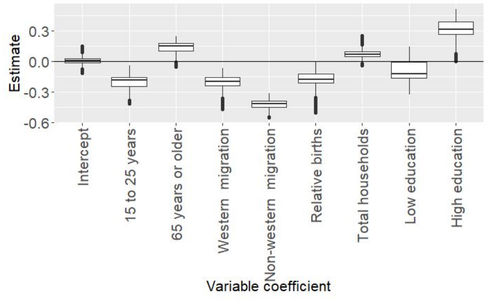


Fig. 2 - Residuals spatial lag model

GWR Global pseudo-R2 = 0.56, AIC = 6250



Municipalities

Spatial lag R2 = 0.59, AIC = 579

Variable	Estimate	SD	p-value
Intercept	-0.0003	0.032	0.99
Western Migration background	**-0.27**	**0.04**	**1e-11**
Total Non-western migration background	**-0.21**	**0.08**	**0.005**
Distance to GP	**+0.13**	**0.04**	**0.001**
Households no children	-0.078	0.05	0.09
Low education	**-0.13**	**0.06**	**0.03**
High education	**+0.27**	**0.06**	**2e-05**
Monthly church visit	**-0.18**	**0.07**	**0.007**
Reformed church	**-0.29**	**0.06**	**1e-06**
Islam	+0.01	0.06	0.8
'Different' religion	**-0.19**	**0.04**	**2e-05**

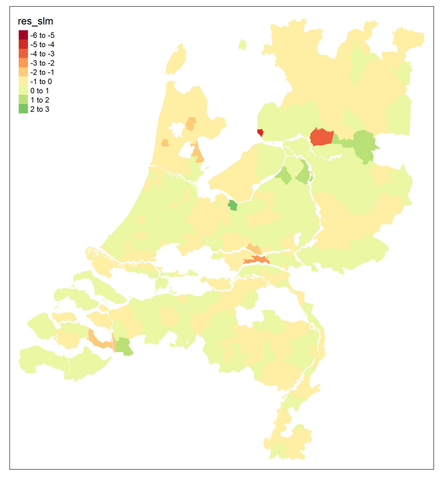
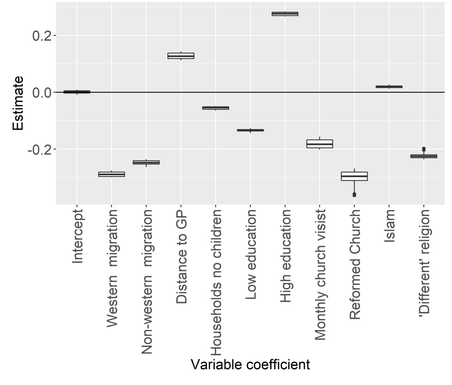


Fig. 3 - Residuals spatial lag model

GWR Global pseudo-R2 = 0.66, AIC = 570



Data

Table 1 - Data used in model

Topic	Level	Year	Authority
Vaccine coverage	Neighbourhood Municipality	2022	RIVM
Demographics	Neighbourhood Municipality	2020	CBS
Education	Neighbourhood Municipality	2020	CBS
Religion	Municipality	2010-2014	CBS

Discussion

Municipalities in the Netherlands change over time due to fusions of municipalities. The datasets are from different points in time, resulting in missing values in the data or results in incorrect spatial joining of data - a core issue.

A source of bias in the model is aggregation bias. This is a cause of concern in the model on the municipality level as the data is aggregated to a bigger region averaging out variance of its neighbourhoods. For further improvements it would be much better to have religion data on a neighbourhood level and to have data from the same time period.

The predictor, distance to GP is significant. It was tested if distance to GP was a confounder for population density since that is what it correlates with, but it was not since population density was not found to be a significant predictor.

Another limitation to our models is the absence of a validation set. However, splitting the data in train and test data would greatly influence the spatial correlations this study was interested in finding. A reasonable validation would therefore be running the model to predict vaccine uptake during a future pandemic.

Comparison of the AIC and R2 of the different models on same level shows that for neighbourhoods spatial lag performs better and on municipalities GWR performs better. This can be explained by the fact that neighbourhoods have a lot more variance and this also reflects on the variance of the coefficient estimates of the GWR model for the neighbourhood level.

Conclusion

In this paper multiple socio-demographic factors such as gender, age groups, ethnic groups, household types, education and religious affiliation were were investigated using different spatial models at both neighbourhood and municipality level. It was found that the fixed GWR was the most effective model at municipality level and the spatial lag model most effective at neighbourhood level. Distance to GP was indicative of whether an individual received a vaccine or not. At municipality level, religious affiliation showed negative impact.

Bibliography

1. Melinda Mills and David Salisbury. The challenges of distributing covid-19 vaccinations. EClinicalMedicine, 31:100674, 12 2020.