GeoHealth

Supporting Information for

Spatial Effects of Livestock Farming on Human Infections with Shiga Toxin-Producing Escherichia coli O157 in Small But Densely Populated Regions: The Case of the Netherlands

A.C. Mulder[1], J. van de Kassteele[1], D. Heederik[2], R. Pijnacker[1], L. Mughini-Gras[1,2,†] and E. Franz[1,†]

[1]Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Postbus 1, 3720 BA Bilthoven, the Netherlands.

[2]Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, Utrecht, the Netherlands.

Corresponding author: Annemieke C. Mulder (annemieke.mulder@rivm.nl)

[†]These authors contributed equally to this article and share last authorship.

Contents of this file

- Text S1 to S4
- Figures S1 to S3
- Tables S1 to S2

Introduction

This file mainly contains supporting information supporting the Materials and Methods section of the article, including the following items:

- An explanation of the population-weighted interpolation (**Text S1**)
- A figure showing the changing four-digit postal code regions of the Netherlands over the years (**Figure S1**)
- An overview of the R packages used (**Table S1**)
- An explanation of the interpretation of the rate ratios used (RR) (**Text S2**)
- A figure visualizing this interpretation (**Figure S2**)
- An explanation of how the population attributable fraction was calculated (PAF) (**Text S3**)
- An explanation of the different spatial scales of the Netherlands compared to the NUTS classification system (**Text S4**)
- A figure showing those different spatial scales (**Figure S3**)
- An overview of the univariable spatial regression results (**Table S2**)

Text S1. Population weighted interpolation

The population weighted interpolation was carried out as follows: first, we made an intersection between the four-digit postal code regions and the six-digit postal code points. Next, the four-digit postal code region data (both the STEC O157 cases and the population numbers by age category and gender) were redistributed over the six-digit postal code points, proportional to the number of inhabitants for these six-digit postal code point locations. Then, an intersection was made between the six-digit postal code points and the hexagonal grid. Finally, the redistributed data over the six-digit postal code points were allocated to each hexagon.

Some four-digit postal code regions could not be redistributed, because no six-digit postal code points could be assigned to it. In that case, the nearest six-digit postal code point location was used. Similarly, when a six-digit postal code point could not be assigned to a hexagon, the nearest hexagon was used. The redistribution from the four-digit postal code regions to six-digit postal code points to the hexagonal grid could be done very efficiently by sparse matrix multiplications. For each age category and gender stratum, the same redistribution matrix was used.

Text S2. Rate ratio (RR)

In this study, the exposure measure x can get the value zero. Therefore, the explanatory variable was transformed using the $log_2(x+1)$ transformation. Resulting in the following Poisson regression with log-link function formula:

$$\log(\mu) = \beta_0 + \beta_1 log_2(x+1)$$

By taking the inverse link-function of this, using the exponential function e^x, we obtained:

$$\begin{aligned} \mu &= e^{\beta_0+\beta_1(x+1)} \\ &= e^{\beta_0} e^{\beta_1(x+1)} \end{aligned}$$

The RR_{21} for an exposure at $log_2(x_2+1)$ relative to $log_2(x_1+1)$ then is:

$$RR_{21} = e^{\beta_1 log_2(\frac{x_2+1}{x_1+1})}$$

If $x+1$ grows with a factor two, the rate increases with a factor $RR = e^{\beta_1}$. Fortunately, not much changes when x is large relative to one, as the following applies:

$$\frac{x_2+1}{x_1+1} \approx \frac{x_2}{x_1}$$

This leads to the same "easier" interpretation of the rate ratio as when using a $\log(x)$ transformation: if x increases with a factor two, the incidence rate increases with a factor $RR = e^{\beta_1}$. But what is "large" ? Do we make a big mistake with this approximation? We visualized this in Figure S2. In this figure, x_1 increases from one towards 1,000 and the factor two was chosen as ratio between x_2 and x_1, thus $x_2 = 2x_1$. The x-axis was transformed into a log_{10} scale to make the effect of large values of x_1 on the factor more clear. The constant value of two is what we would have at $\frac{x_2}{x_1} = 2$. The red line is this factor when we add one to x. As Figure S2 shows, this approximation is pretty good when values of x_1 are approximately above 100. This indicates that the "easier" interpretation of the rate ratio can be used.

In summary, if the $log_2(x_2+1)$ is used as explanatory variable in Poisson regression with log-link function, then the interpretation of the rate ratio (RR) is as follows: if x increases with a factor two, then the incidence rate increases with a factor $RR = e^{\beta_1}$, provided that x is large enough, approximately >100. When x is smaller, this factor is less than two for the same RR, but the significance stays the same.

Text S3. Population attributable fraction (PAF)

The PAF is calculated as follows:

$$PAF = \left(\frac{i(E) - i(0)}{i(E)}\right) * 100$$

Here $i(E)$ is the predicted incidence in the exposed population (using the regression model and its estimated coefficients as is) and $i(0)$ is the predicted incidence in the unexposed population (using the same regression model and estimated coefficients, but where the exposure of the risk factor is set to zero). Both predictions can be done simultaneously by augmenting the original dataset, where in the augmented records the exposure of the risk factor is set to zero and the outcome is set to missing. For each group (exposed and non-exposed), the total incidences are calculated as the sum of the individual records.

Text S4. Spatial scales of the Netherlands

To divide the economic territory of the EU, a hierarchical system was developed. This system is called the NUTS classification (Nomenclature of territorial units for statistics) (European Commission - Eurostat, 2019). It contains three levels:

- NUTS 1: major socio-economic regions
- NUTS 2: basic regions for the application of regional policies
- NUTS 3: small regions for specific diagnosis.

The current NUTS 2016 classification is valid from 1 January 2018 and lists 104 regions at NUTS 1, 281 regions at NUTS 2 and 1348 regions at NUTS 3 level (European Commission - Eurostat, 2019). In the Netherlands, the NUTS 1 regions consist of four areas: North of the Netherlands, East of the Netherlands, West of the Netherlands and South of the Netherlands. The NUTS 2 regions are the Dutch provinces (**Figure S3** - a) and the NUTS 3 regions are 40 COROP regions, which consist of a combination of several municipalities of a province. Thus, the municipalities in the Netherlands (~ 90 km2, **Figure S3** - b) are smaller than those NUTS 3 regions and the four-digit postal code regions of the Netherlands (~ 10 km2, **Figure S3** - c) are even smaller than those municipalities. The six-digit postal code point locations of the Netherlands give information about specific locations at street level.

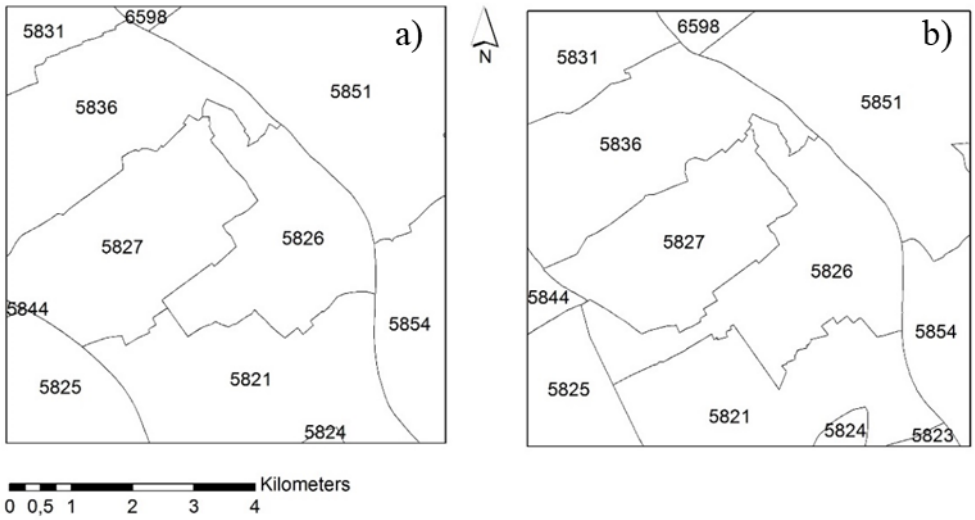

Figure S1. Example of changing four-digit postal code regions of the Netherlands over the years; a) 2009 compared to b) 2016.

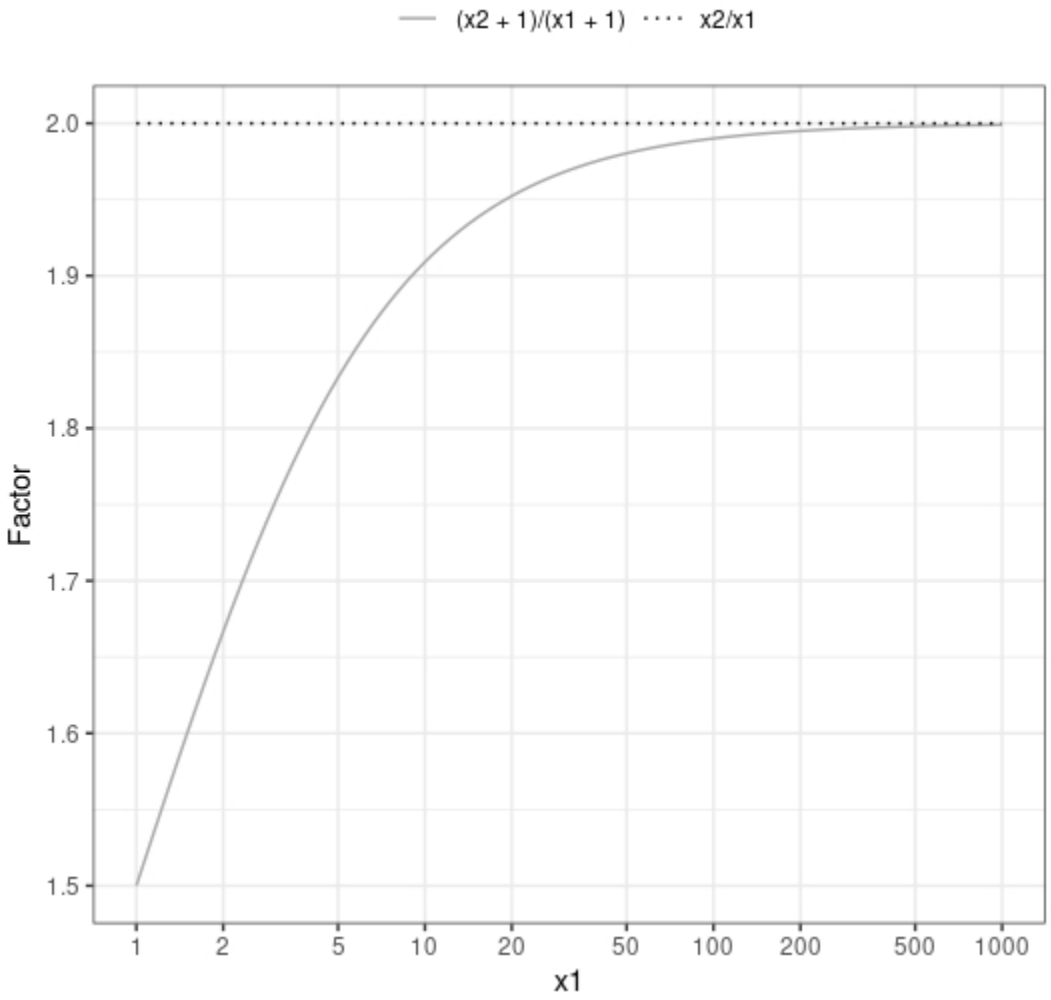

Figure S2. Visualization of interpretation rate ratio (RR) for an exposure measure $\boldsymbol{x_1}$.

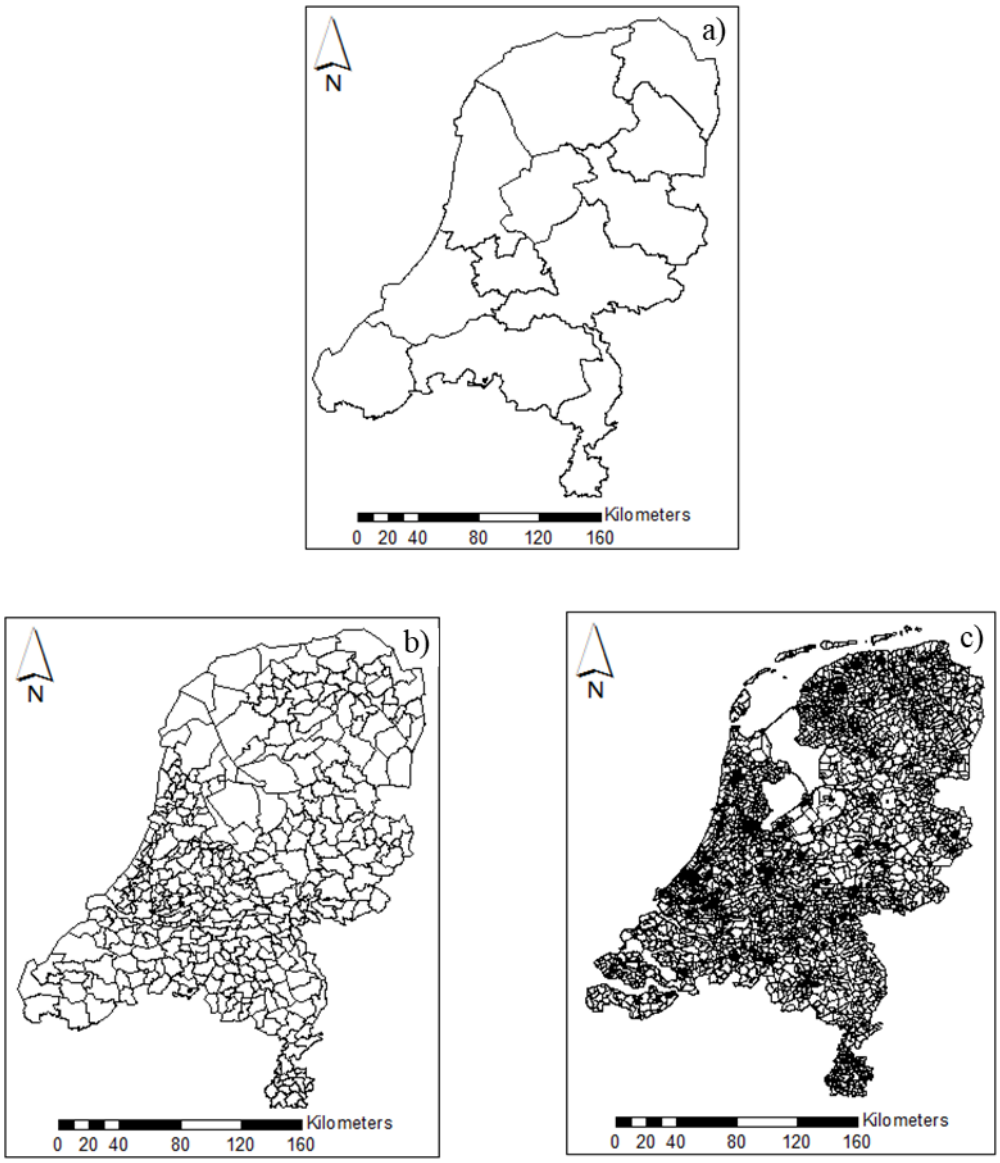

Figure S3. The different administrative boundaries and spatial scales of the Netherlands. a) Provinces (NUTS 2 regions), b) Municipalities, c) Four-digit postal code regions.

Package/function	Version	Reference
cbsodataR	0.3.5	(De Jonge & Houweling, 2019)
dplyr	0.8.3	(Wickham, Francois, Henry, & Müller, 2019)
INLA	18.07.12	(H. Rue, 2019)
lubridate	1.7.4	(Grolemund & Wickham, 2011)
Matrix	1.2-17	(Bates, Maechler, Davis, Oehlschlägel, & Riedy, 2019)
Parallel	3.6.0	(R-Core, 2017)
RANN	1.2.6	(Arya, Mount, Kemp, & Jefferis, 2015)
RColorBrewer	1.1-2	(Neuwirth, 2015)
readxl	1.3.1	(H. Wickham, J. Bryan, et al., 2019)
rgdal	1.4-4	(Keitt, 2010)
sf	0.7-7	(Pebesma, Bivand, Racine, et al., 2019)
sp	1.3-1	(Pebesma, Bivand, Rowlingson, et al., 2019)
spdep	1.1-2	(Bivand et al., 2019)
st_make_grid	-	(Pebesma, 2019)
stringr	1.4.0	(Wickham, 2019)
tidyr	1.0.0	(Wickham, Henry, & Rstudio, 2019)
tidyverse	1.3.0	(Hadley Wickham et al., 2019)

Table S1. An overview of the R packages and functions used, including version numbers and references.

	Hexagon 90 km2			Hexagon 50 km2			Hexagon 25 km2			Hexagon 10 km2		
Variable	**P-value**	**RR**	**95% CI**	**P-value**	**RR**	**95% CI**	**P-value**	**RR**	**95% CI**	**P-value**	**RR**	**95% CI**
Period of infection												
Winter						Reference category						
Summer	**<0.001**	**3.43**	**2.76-4.31**	**<0.001**	**3.43**	**2.76-4.31**	**<0.001**	**3.43**	**2.76-4.31**	**<0.001**	**3.43**	**2.76-4.31**
Gender												
Males						Reference category						
Females	**<0.001**	**1.60**	**1.32-1.95**	**<0.001**	**1.60**	**1.32-1.95**	**<0.001**	**1.60**	**1.32-1.95**	**<0.001**	**1.60**	**1.32-1.95**
Age category (years)												
0-4	**<0.001**	**3.70**	**2.75-4.95**	**<0.001**	**3.71**	**2.76-4.96**	**<0.001**	**3.70**	**2.75-4.95**	**<0.001**	**3.70**	**2.75-4.95**
5-9	**<0.001**	**2.16**	**1.52-3.03**	**<0.001**	**2.17**	**1.53-3.04**	**<0.001**	**2.17**	**1.52-3.04**	**<0.001**	**2.17**	**1.52-3.03**
10-49	0.30	1.13	0.90-1.41	0.30	1.13	0.90-1.41	0.31	1.12	0.90-1.41	0.31	1.12	0.90-1.41
≥ 50						Reference category						
Type of animal [a]												
Small ruminants	**<0.01**	**1.12**	**1.04-1.20**	**<0.001**	**1.12**	**1.05-1.19**	**<0.01**	**1.08**	**1.03-1.14**	0.05	1.04	1.00-1.09
Cattle	0.07	1.07	0.99-1.14	0.13	1.05	0.99-1.11	0.10	1.04	0.99-1.09	**0.04**	**1.04**	**1.00-1.08**
Poultry	0.17	1.02	0.99-1.05	0.36	1.01	0.99-1.04	0.23	1.01	0.99-1.04	0.26	1.01	0.99-1.03
Pigs	0.11	1.03	0.99-1.08	**0.03**	**1.04**	**1.00-1.08**	0.10	1.03	0.99-1.06	**0.01**	**1.04**	**1.01-1.07**

[a] *Population weighted number of animals*

Table S2. Univariable spatial analyses results for different hexagonal areas (90, 50, 25 and 10 km2).