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

Annemieke Christine Mulder[1], Jan van de Kassteele[1], Dick Heederik[2], Roan Pijnacker[1], Lapo Mughini-Gras[3], and Eelco Franz[1]

[1]Dutch Institute for Public Health and the Environment (RIVM)
[2]Utrecht University
[3]Dutch Institute for Public Health and the Environment (RIVM) & Utrecht University

November 24, 2022

Abstract

The role of environmental transmission of typically foodborne pathogens like Shiga toxin-producing (STEC) O157 is increasingly recognized. To gain more insights, we assessed the spatial association between sporadic STEC O157 human infections and the exposure to livestock (i.e. small ruminants, cattle, poultry, and pigs) in a densely populated country: the Netherlands. This was done for the years 2007-2016, using a state-of-the-art spatial analysis method in which hexagonal areas with different sizes (90, 50, 25 and 10 km2) were used in combination with a novel probability of exposure metric: the population-weighted number of animals per hexagon. To identify risk factors for STEC O157 infections and their population attributable fraction (PAF), a spatial regression model was fitted using integrated nested Laplace approximation (INLA). Living in hexagonal areas of 25, 50 and 90 km2 with twice as much population-weighted small ruminants was associated with an increase of the incidence rate of human STEC O157 infections in summer (RR of 1.09 [95%CI;1.01-1.17], RR of 1.17 [95%CI;1.07-1.28] and RR of 1.13 [95%CI;1.01-1.26]), with a PAF of 49% (95%CI;8-72%). Results indicate a potential transmission of STEC O157 from small ruminants to humans via the environment. However, the underlying mechanisms warrant further investigation and could offer new targets for control. Furthermore, the newly proposed exposure metric has potential to improve existing spatial modelling studies on infectious diseases related to livestock exposure, especially in densely populated countries like the Netherlands.

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

4

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

7

8 [1]Centre for Infectious Disease Control, National Institute for Public Health and the Environment
9 (RIVM), Postbus 1, 3720 BA Bilthoven, the Netherlands.

10 [2]Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology,
11 Utrecht University, Utrecht, the Netherlands.

12

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

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

15

16 **Key Points:**

17 • Results indicate a potential transmission of STEC O157 from small ruminants, and not
18   cattle, to humans via the environment.

19 • However, the underlying mechanisms warrant further investigation and corresponding
20   results could offer new targets for control.

21 • The newly proposed exposure metric has potential to improve similar spatial modelling
22   studies, especially in densely populated countries.
23

24 **Abstract**

25 The role of environmental transmission of typically foodborne pathogens like Shiga toxin-
26 producing *Escherichia coli* (STEC) O157 is increasingly recognized. To gain more insights, we
27 assessed the spatial association between sporadic STEC O157 human infections and the exposure
28 to livestock (i.e. small ruminants, cattle, poultry, and pigs) in a densely populated country: the
29 Netherlands. This was done for the years 2007-2016, using a state-of-the-art spatial analysis
30 method in which hexagonal areas with different sizes (90, 50, 25 and 10 km2) were used in
31 combination with a novel probability of exposure metric: the population-weighted number of
32 animals per hexagon. To identify risk factors for STEC O157 infections and their population
33 attributable fraction (PAF), a spatial regression model was fitted using integrated nested Laplace
34 approximation (INLA). Living in hexagonal areas of 25, 50 and 90 km2 with twice as much
35 population-weighted small ruminants was associated with an increase of the incidence rate of
36 human STEC O157 infections in summer (RR of 1.09 [95%CI;1.01-1.17], RR of 1.17
37 [95%CI;1.07-1.28] and RR of 1.13 [95%CI;1.01-1.26]), with a PAF of 49% (95%CI;8-72%).
38 Results indicate a potential transmission of STEC O157 from small ruminants to humans via the
39 environment. However, the underlying mechanisms warrant further investigation and could offer
40 new targets for control. Furthermore, the newly proposed exposure metric has potential to
41 improve existing spatial modelling studies on infectious diseases related to livestock exposure,
42 especially in densely populated countries like the Netherlands.

43

44 **Plain Language Summary**

45 I will add this summary during the revision stage

46

47 **Key words:** STEC O157, livestock, small ruminants, population-weighted number of animals,
48 environmental transmission

49

1 Introduction

51   Food is generally considered to be the most important route of transmission for Shiga
52 toxin-producing *Escherichia coli* (STEC) O157 (N. J. C. Strachan et al., 2001). However, a
53 growing body of evidence suggests that non-foodborne transmission pathways, such as those
54 mediated by the environment, may be important as well (Berry et al., 2015; Elson et al., 2018;
55 Franz et al., 2018; Friesema et al., 2011; ÓHaiseadha et al., 2017; Norval J. C. Strachan et al.,
56 2006). A recent source attribution modelling study based on STEC serotyping data revealed that
57 domestic ruminants (cattle, sheep and goats) are important sources of human STEC O157
58 infections, accounting for approximately three-quarters of reported human STEC infections in
59 the Netherlands (Mughini-Gras et al., 2018). This emphasizes the need for both direct and
60 indirect exposure to different types of livestock to be considered as possible transmission routes
61 for STEC O157.
62   STEC is a bacterial zoonotic agent associated with human disease with varying clinical
63 manifestations, including diarrhea, haemorrhagic colitis and (occasionally fatal) haemolytic
64 uremic syndrome (HUS), a leading cause of acute renal failure among children (Elson et al.,
65 2018; Franz et al., 2018; Mughini-Gras et al., 2018). Human STEC infections is the third most

66   commonly reported zoonosis in the European Union (EU), with an annual number of laboratory-
67   confirmed STEC infections increasing from 5,901 in 2014 to 8,161 in 2018 (European Food
68   Safety Authority & European Centre for Disease Prevention and Control, 2019). This, in
69   combination with its high virulence and outbreak potential, makes STEC of significant public
70   health concern. Although there are more than a hundred STEC serotypes and their importance is
71   increasingly recognized, STEC O157 is the most important serotype in terms of incidence and
72   clinical significance (Mughini-Gras et al., 2018). In the Netherlands, STEC is a notifiable
73   disease, with an annual incidence between 2 and 7 cases per 100.000 inhabitants (European
74   Centre for Disease Prevention and Control, 2019). The vast majority of cases in the Netherlands
75   are considered sporadic, as outbreaks rarely occur (Franz et al., 2018).
76   Potential sources of human STEC infection are mainly animals capable of maintaining
77   STEC colonization in absence of continuous exposure to STEC from other sources (i.e. the so-
78   called reservoirs or amplifying hosts, mainly cattle and sheep). But also those that are frequently
79   exposed to STEC from the environment, like birds and other wild animals (Mughini-Gras et al.,
80   2018; Norval J. C. Strachan et al., 2006). According to a recent source attribution study, cattle is
81   the primary source of human STEC O157 infection in the Netherlands, followed by small
82   ruminants (sheep and goats) (Mughini-Gras et al., 2018). These animals can shed high quantities
83   (>105/g) of STEC O157, that subsequently are able to survive for extended periods of time
84   (Chase-Topping et al., 2008; Franz et al., 2014; N. J. C. Strachan et al., 2001). This implies that
85   there is a significant risk of STEC O157 infection linked to environment-mediated transmission
86   to humans (Elson et al., 2018; N. J. C. Strachan et al., 2001).
87   The Netherlands is one of the world's most densely populated countries, with over 500
88   inhabitants per km2 and a remarkably high concentration of intensive livestock farms as well.
89   The presence of livestock in close proximity to residential areas has arisen questions about the
90   associated public health implications (Smit & Heederik, 2017). Since STEC O157 can
91   potentially be contracted from the soil and water environment, and may be spread through the air
92   after periods of drought in the vicinity of its animal reservoirs, it is conceivable that human
93   STEC O157 incidence in the Netherlands might higher in areas with increased livestock density
94   as well, such as in rural vs. urban areas as shown elsewhere (Berry et al., 2015; Norval J. C.
95   Strachan et al., 2006; N. J. C. Strachan et al., 2001). This could be tested with different methods,
96   of which examples are: (i) spatial regression analysis to determine the probability of exposure
97   (Elson et al., 2018; Friesema et al., 2011; ÓHaiseadha et al., 2017) or (ii) classical case-control
98   studies including relevant spatial variables to determine the importance of particular types of
99   exposure (e.g. number of animals/km2) (de Rooij et al., 2019).
100   As a spatial regression analyses requires less resources, in terms of data needs and
101   financial support, it can be a preferred way of exploring new ideas. However, only a few studies
102   exist that focus on the spatial association between human STEC O157 infections and the
103   probability of exposure to livestock by means of spatial regression analysis (Elson et al., 2018;
104   Friesema et al., 2011; ÓHaiseadha et al., 2017). Most of those studies only include one domestic
105   ruminant species (cattle or sheep or goat) in the analysis (Friesema et al., 2011; ÓHaiseadha et
106   al., 2017), while ignoring other reservoirs that may affect the outcome of those studies. This is
107   especially important in countries like the Netherlands where high numbers of different types of
108   livestock are present on relatively small geographical scales (Smit & Heederik, 2017). Moreover,
109   the probability of exposure in those studies is strictly defined by the number of animals in a
110   given area, while the probability of exposure on a population level is not only determined by the
111   number of animals in a certain area but also by the number of residents living in that area (Elson

112 et al., 2018; Friesema et al., 2011; Hallisey et al., 2017; Mulder et al., 2016; ÓHaiseadha et al.,
113 2017).
114 Therefore, the aim of this study was to assess the spatial association between sporadic
115 human STEC O157 infections and the combined exposures to livestock (cattle, goat, sheep,
116 poultry and pigs) in the Netherlands, using different state-of-the-art methods that include
117 population-weighted numbers of animals in the calculation of the probability of exposure to
118 livestock.

2 Materials and Methods

120 This study consisted of several parts. First, national surveillance data on notified STEC
121 O157 cases in the Netherlands' general population was gathered together with livestock data
122 (exact locations of registered farms and number of animals therein, per species). Subsequently,
123 the data were transposed into a study-defined spatial division of the Netherlands and we
124 developed a metric for the probability of exposure of the human population to each livestock
125 species that not only includes the number of animals in a certain area, but also the corresponding
126 population number. The last steps involved the spatial regression analysis and calculation of the
127 population attributable fraction (PAF). We used the statistical software environment R (version
128 3.6.0) (RCT, 2015) and several R packages and functions for data processing and analysis (Arya
129 et al., 2015; Bates et al., 2019; Bivand et al., 2019; De Jonge & Houweling, 2019; Grolemund &
130 Wickham, 2011; Keitt, 2010; Neuwirth, 2015; Pebesma, 2019; Pebesma, Bivand, Racine, et al.,
131 2019; Pebesma, Bivand, Rowlingson, et al., 2019; R-Core, 2017; Rue, 2019; Wickham, 2019;
132 Wickham, Averick, et al., 2019; Wickham, Bryan, et al., 2019; Wickham, Francois, et al., 2019;
133 Wickham, Henry, et al., 2019). An overview is provided in supporting information Table S1. The
134 used R scripts can be found at: https://github.com/mulderac91/R-STECO157-spatialanalysis

135 2.1 Hexagonal grid and population-weighted interpolation

136 Hexagons are more suitable than rectangular grids in particular applications of ecological
137 modelling, e.g. connectivity and movement paths (Birch et al., 2007). They have the advantage
138 that the nearest neighborhood in a hexagonal grid is simpler and less ambiguous, because each
139 hexagon has exactly six adjacent hexagons which are in a symmetrically equivalent position.
140 Therefore, there is no need for a setting for the relative weighting of diagonal interactions in a
141 nearest neighborhood analysis, as is the case for rectangular grids (Birch et al., 2007; Birch et al.,
142 2000). Furthermore, the grid is fixed over time (Birch et al., 2007). The latter is a solution for the
143 problem of change of, in this case postal code boundaries over time (supporting information
144 Figure S1). Therefore, the Netherlands was divided in a fixed hexagonal grid (Figure 1 - a). To
145 assess consistency of results and reduce the risk of ecological fallacy, we performed the analyses
146 for hexagonal areas with four different sizes: 10 km2 (approximately the average area of a four-
147 digit postal code region in the Netherlands), 25 km2, 50 km2 and 90 km2 (approximately the
148 average area of a municipality in the Netherlands) (Shafran-Nathan et al., 2017).

149 In order to perform the spatial regression analyses on the hexagonal grid, the spatial data
150 needed to be transformed from one regional division to the other (Arsenault et al., 2013). For this
151 purpose, we used population-weighted interpolation. This approach has the advantage over areal
152 weighted interpolation that it can more accurately estimate the population demographics in
153 transforming small counts by four-digit postal code regions to aggregated counts for large, non-

154 standard study zones (hexagons) (Hallisey et al., 2017). A detailed explanation of this approach
155 can be found in supporting information Text S1.

156   2.2 Population-weighted number of animals

157   Existing studies have used animals/km2 to derive the probability of exposure to be able to
158 study the association between STEC O157 infections and livestock densities (Figure 1 - b)
159 (Elson et al., 2018; Friesema et al., 2011; ÓHaiseadha et al., 2017). Yet, the probability of
160 exposure is not only determined by the number of animals in a certain area, but also by the
161 number and residential addresses of people living in that area and the number of animals in the
162 neighboring areas. For this purpose, we created a new probability of exposure metric: the
163 population-weighted number of animals (Hallisey et al., 2017) (Figure 1-c, d and e).
164   The metric is constructed as follows. When zooming into one hexagon within the
165 hexagonal grid, the locations of several six-digit postal code points are shown (Figure 1-c).
166 Those six-digit postal code points include information about the population numbers at that
167 specific location (Figure 1-c). Around these point locations, buffers with a radius of 1 km are
168 constructed (Figure 1-d). Farms located within these buffers, also outside the specified hexagon,
169 are included (Figure 1-e). The point locations of the farms contain information about the number
170 of animals (Figure 1-e). See Figure 1-c, d and e as an example. Within the hexagon, we have five
171 six-digit postal code point locations, each with its own population numbers: 100, 1,000, 10, 5 and
172 1. We have three farms, each with its own number of animals: A, B and C. The 100 and the 10
173 individuals on the first and second six-digit postal code point locations are exposed to A animals.
174 The 1,000 individuals in the third six-digit postal code point location are not exposed. The 5
175 individuals in the fourth six-digit postal code point location are exposed to B animals. The only
176 individual in the fifth six-digit postal code point location is exposed to C animals, but from a
177 farm outside the hexagon. The total exposure in this hexagon is then the population-weighted
178 sum of the number of animals, which can be calculated as follows:

179

180 $$Population\ weighted\ animal\ number\ = \frac{(100\ x\ A + 10\ x\ A + 1{,}000\ x\ 0 + 5\ x\ B + 1\ x\ C)}{(100 + 10 + 1{,}000 + 5 + 1)}$$

181

182   This was done for each hexagon and for each year, taking into account the number of
183 animals and the changing population numbers. In the end, the data were aggregated over the
184 years, resulting in one hexagon-specific exposure metric.

185
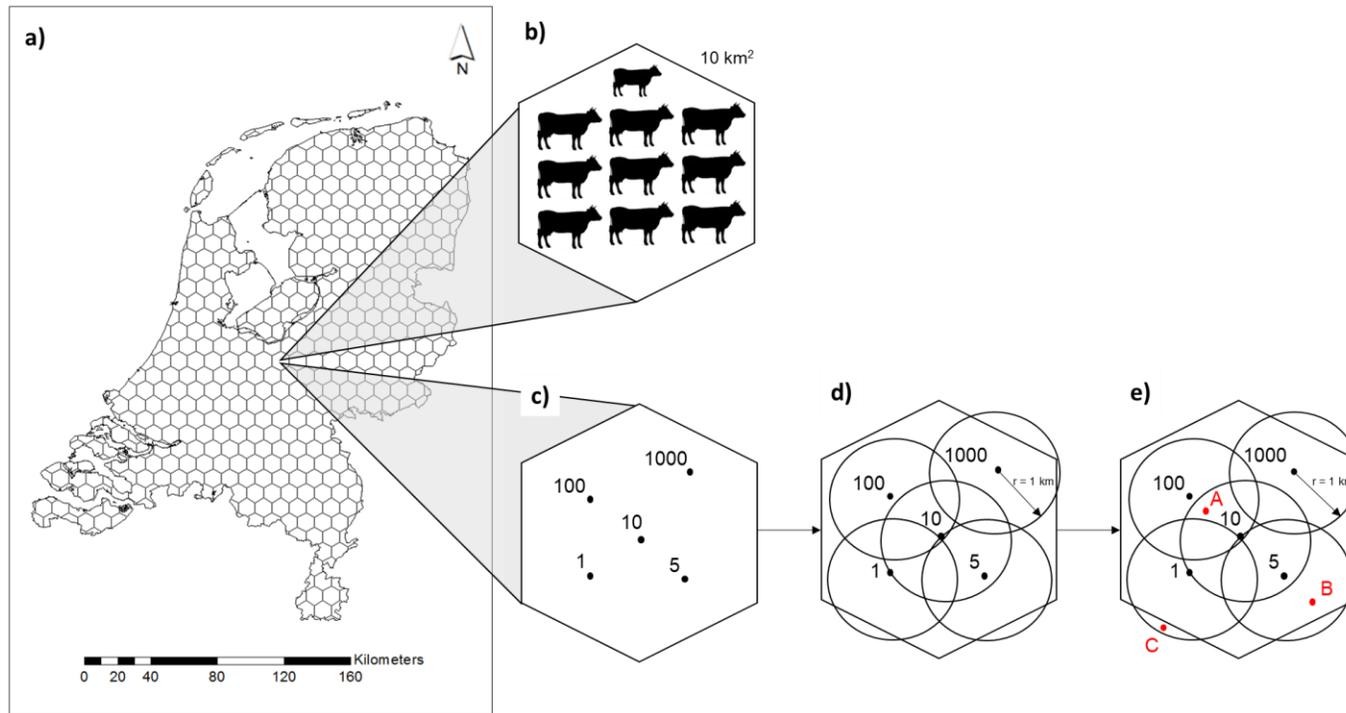
Figure 1.

187 Explanation of the calculation of the old and the new probability of exposure measures. **a)** Hexagonal grid of the Netherlands. **b)** The
188 old probability of exposure measure in a hexagonal grid cell: the number of animals per km2. In this figure: 10 cows per 10 km2, thus
189 1 cow/km2. Pictures **c**, **d** and **e** visualize the calculation of the new probability of exposure: the population-weighted animal number.
190 **c)** A hexagonal grid cell, including the six-digit postal code point locations within this cell and their corresponding population
191 numbers. **d)** The buffers with a radius r of 1 km surrounding the six-digit postal code point locations within the hexagonal grid cell. **e)**
192 The hexagonal grid cell including all the information of Figure 1 - d. Here, the point locations of the farms of a certain type of animal
193 are added, which also include information about the specific number of animals. This gives the information that is needed to know
194 which six-digit postal code points (and thus which population numbers) are influenced by which farm(s) and the corresponding animal
195 numbers. With this information and the formula given in Section 2.2, the new probability of exposure can be calculated and
196 aggregated per hexagon.

197   2.3 Spatial risk factor analysis

198

199   A Poisson regression model with log-link function was used to assess the associations
200 between human STEC O157 infections and the population-weighted number of animals for
201 cattle, pigs, poultry, and small ruminants (goats and sheep). The model described the number of
202 human cases as a function of person-years (the population denominator for each hexagon), age
203 category (0-4, 5-9, 10-49 and ≥50 years old), gender (male or female), period of infection
204 (spring/summer: May-October, autumn/winter: November-April) and the different population-
205 weighted number of animals (Friesema et al., 2011). Because the population-weighted number of
206 animals x could be zero, we applied a $log_2(x + 1)$ transformation. Furthermore, several studies
207 have shown a higher risk for human STEC O157 infection in summer (Friesema et al., 2011).
208 Therefore, we performed a stratified analysis based on the period of infection. These variables
209 entered the model as the fixed effect terms.
210   It is possible that there is additional variation due to unknown spatially varying risk factors.
211 To account for this, two random-effect terms were added to the model. The first random-effect
212 accounted for the spatially structured variation. This variation represented the possible effect of a
213 common unobserved risk factor that led to neighboring hexagons being more alike. This term
214 was modelled by the intrinsic Conditional Autoregressive Model (CAR) (Besag et al., 1991). The
215 second random-effect term represented the unstructured variation. This variation consisted of
216 possible unobserved variation within hexagons, which was modelled by independent and
217 identically distributed (IID) Gaussian noise (Lawson, 2013).
218   The spatial regression model was fitted using the integrated nested Laplace approximation
219 technique (INLA) (Rue et al., 2019). For further details we refer to Friesema et al. (2011). Rate
220 ratios (RRs) were calculated from the coefficients of the fixed effects. As the population-
221 weighted animal numbers were transformed, the interpretation of those RRs is as follows: if x
222 increases with a factor two, then the incidence rate increases with a factor $RR = e^{\beta_1}$, provided
223 that x is large enough, approximately >100. When x is smaller, this factor is less than two for the
224 same RR, but the significance stays the same. Supporting information Text S2 and Figure S2
225 show a more detailed explanation of this interpretation.
226   In addition, the population attributable fraction (PAF) and its 95% confidence interval were
227 calculated for the risk factors found (supporting information Text S3). Confidence intervals were
228 obtained by Monte Carlo simulation, using the INLA posterior sampling function with 10,000
229 samples.

230 **3 Data**

231   3.1 Case data

232   Since 1999, it is obligatory for diagnostic laboratories in the Netherlands to notify confirmed
233 human STEC infections to the Municipal Health Services (MHSs) (Friesema et al., 2011). The
234 MHSs reports each laboratory-confirmed case to the national surveillance database at the Dutch
235 National Institute for Public Health and the Environment (RIVM) (Friesema et al., 2011).
236 Furthermore, laboratories are asked (but not obliged) to send STEC isolates to the RIVM for
237 confirmation and further typing for national surveillance purposes (Friesema et al., 2017;
238 Friesema et al., 2011).

239   In this study, a case was defined as an individual with confirmed STEC O157 infection (by

240   the RIVM) during the period 2007-2016. Cases were excluded when they were part of an

241   (inter)national foodborne outbreak, travelled abroad in the week before onset of illness, or when

242   the residential address (postal code) was unknown. A detailed explanation of the different spatial

243   scales (province, municipality and postal code) of the Netherlands and a comparison with the

244   European NUTS classification system is given in supporting information Text S4 and Figure S3

245   (European Commission - Eurostat, 2019). Those data are protected by Dutch privacy regulations

246   and the Dutch Data Protection Authority (Dutch Data Protection Authority, 2020a, 2020b).

247   3.2 Livestock data

248   Livestock data for 2012 was obtained from the Department of Service Arrangements of the

249   Dutch Ministry of Agriculture, Nature and Food Quality. These data are collected yearly,

250   requesting all farmers to report the number of animals reared (CBS, 2019b; RVO, 2019). In our

251   study, we used the total number of goats, sheep, cattle, poultry and pigs per farm. To derive the

252   total number of small ruminants, the total number of goats and sheep per farm were summed

253   together.

254   3.3 Population data

255   The population data per four-digit postal code region per year is available through Statistics

256   Netherlands (www.statline.nl) and consists of the number of inhabitants in five-year age

257   categories and gender. The data were downloaded from this website for the years 2007-2016

258   (CBS, 2019a). Due to privacy regulations (Dutch Data Protection Authority, 2020b), this

259   information was not available per six-digit postal code point location.

260   3.4 Spatial data

261   The four-digit postal code region shapefiles of the Netherlands were obtained for each year

262   (2007-2016) from the geodata portal of the RIVM. For the period 2007-2008, there were no

263   postal code region shapefiles available. Therefore, the shapefile of 2009 was used for those

264   years. The six-digit postal code point location shapefile of the Netherlands from 2016 was also

265   obtained from the geodata portal of the RIVM. This file included population numbers per six-

266   digit postal code point location.

267   **4 Results**

268   4.1 Descriptive statistics

269   Between 2007 and 2016, 599 cases of STEC O157 infection were reported. In this period,

270   two national outbreaks of STEC O157 were registered in the Netherlands, one in 2007 involving

271   41 cases probably caused by lettuce consumption and linked to an outbreak in Iceland (Friesema

272   et al., 2008) and one in 2009 involving 20 cases caused by contaminated raw meat spread

273   (Greenland et al., 2009). Furthermore, there was a regional outbreak in 2007 involving 7 cases,

274   which reported consumption of raw meat spread and all had bought it at the same regional

275   supermarket chain (Friesema et al., 2011). The cases that were involved in those outbreaks were

276   excluded from the dataset for analysis. Besides, 54 more cases were excluded because

277   information on travel history prior to symptom onset was missing, and 38 cases because there

278   was no data available on geographical location. The remaining 439 cases were included in the

279 analysis, with a median number of 46 cases per year (range 25-63 cases/year, annual incidence
280 1.5-3.8/100,000 inhabitants).
281 Table 1.

282 *Descriptive Statistics of the STEC O157 Cases.*

	STEC O157 cases	
	N	%
Total	439	100
Gender		
Males	167	38
Females	272	62
Age category (years)		
0-4	70	16
5-9	44	10
10-49	200	46
≥ 50	125	28
Period of infection		
Summer	340	77
Winter	99	23

283
284 Of all the cases included, 62% (n = 272) were female, 38% (n = 167) were male (Table
285 1). The highest number of cases (46%) were between 10 and 49 years of age and most were
286 reported in summer (77%). Figure 2 shows that the incidence varies between hexagons and
287 appears to be highest in the northern and eastern regions of the Netherlands. The west and south
288 of the Netherlands show particularly low incidence of STEC O157.
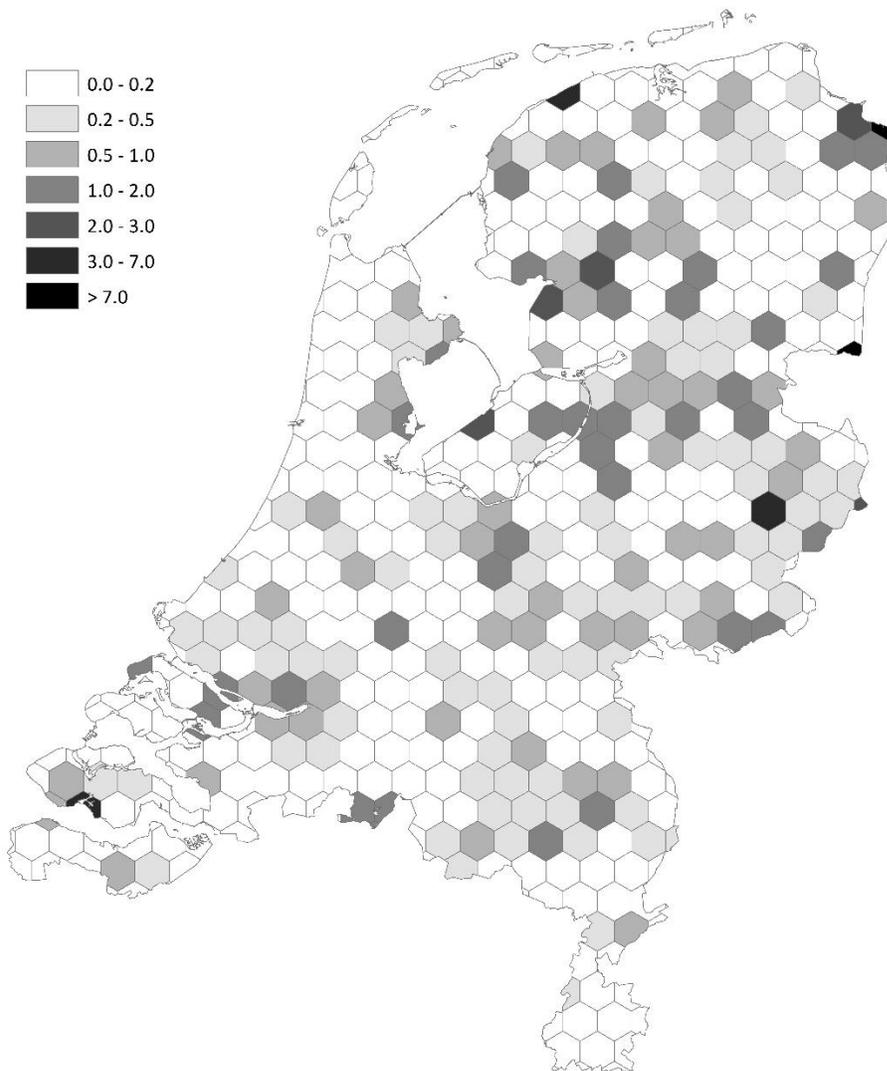
289
290 **Figure 2.**

291 Cumulative incidence rate (x 100.000 person-years) (2007-2016) of STEC O157 infections in the
292 Netherlands.

293   Figure 3 shows the population-weighted number of small ruminants, cattle, poultry and
294 pigs in the Netherlands. The population-weighted number of small ruminants appeared to be
295 highest in the central north of the country, central south of the country and the island of Texel.
296 For cattle, it was highest in the center, central north and central south of the country and for
297 poultry it was highest in the center, east and south-east (except the region of South-Limburg).
298 Furthermore, the population-weighted number of pigs was highest in the east and south-east
299 (except the region south-Limburg). Visually, the map for small ruminants in Figure 3 seemed to
300 be most comparable with the one for human STEC O157 infections in Figure 2.
301
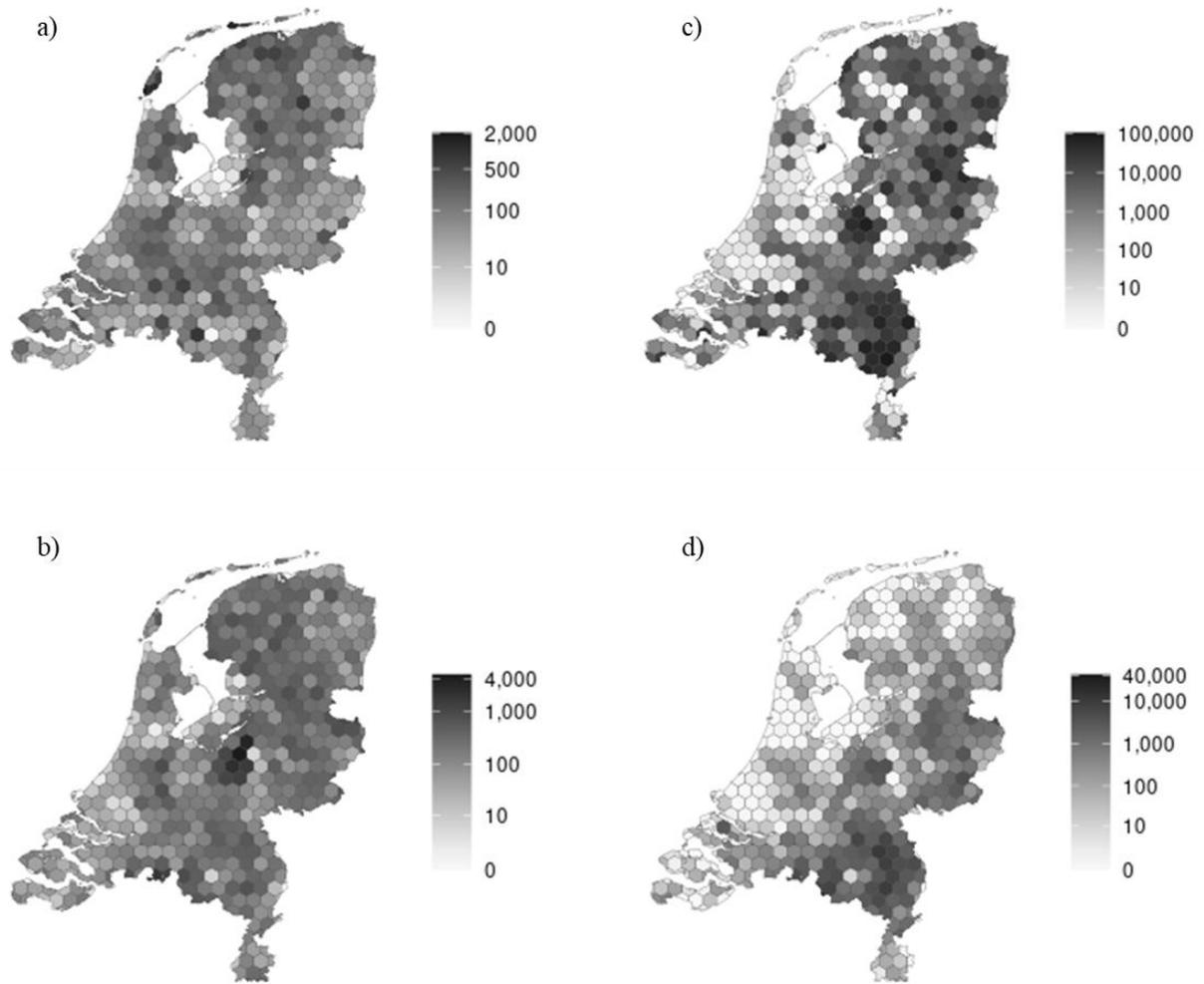
302
303 **Figure 3.**

304 Maps of the population-weighted number of animals in the Netherlands per hexagon (90 km2)
305 for small ruminants (a), cattle (b), poultry (c) and pigs (d) in 2012.

306

307   4.2 Spatial risk factor analysis

308   Results from the multivariable models for the spatial association between STEC O157 and
309 population-weighted number of animals are presented in Table 2 and Table 3, respectively. For
310 the results of the univariable models, see supporting information Table S2.

311 Table 2.

312 *Results of the Multivariable Spatial Analyses for Summer with Different Hexagonal Areas (90, 50, 25 and 10 km2).*

		Hexagon 90 km2			Hexagon 50 km2			Hexagon 25 km2			Hexagon 10 km2		
Period of infection	**Variable**	**P-value**	**RR**	**95% CI**	**P-value**	**RR**	**95% CI**	**P-value**	**RR**	**95% CI**	**P-value**	**RR**	**95% CI**
Summer	Gender												
	Males					*Reference category*							
	Females	**<0.001**	**1.74**	**1.40-2.17**	**<0.001**	**1.74**	**1.40-2.18**	**<0.001**	**1.74**	**1.40-2.17**	**<0.001**	**1.74**	**1.40-2.17**
	Age category (years)												
	0-4	**<0.001**	**4.05**	**2.91-5.59**	**<0.001**	**4.06**	**2.91-5.61**	**<0.001**	**4.06**	**2.91-5.60**	**<0.001**	**4.06**	**2.91-5.60**
	5-9	**<0.001**	**2.01**	**1.32-2.97**	**<0.001**	**2.01**	**1.32-2.98**	**<0.001**	**2.01**	**1.32-2.98**	**<0.001**	**2.01**	**1.32-2.97**
	10-49	0.27	1.15	0.90-1.49	0.27	1.16	0.90-1.49	0.27	1.15	0.90-1.50	0.27	1.16	0.90-1.49
	≥ 50 (ref)					*Reference category*							
	Type of animal [a]												
	Small ruminants	**0.03**	**1.13**	**1.01-1.26**	**<0.001**	**1.17**	**1.07-1.28**	**0.02**	**1.09**	**1.01-1.17**	0.14	1.05	0.99-1.11
	Cattle	0.69	0.97	0.86-1.11	0.20	0.94	0.85-1.03	0.38	0.97	0.89-1.04	0.60	0.98	0.92-1.05
	Poultry	0.50	1.01	0.97-1.06	0.76	0.99	0.96-1.03	0.91	1.00	0.97-1.03	0.96	1.00	0.98-1.03
	Pigs	0.83	1.01	0.94-1.07	0.21	1.04	0.98-1.10	0.28	1.03	0.98-1.08	**0.03**	**1.05**	**1.01-1.09**

[a] *Population-weighted number of animals*

313
314

315   Table 3.

316   *Results of the Multivariable Spatial Analyses for Winter with Different Hexagonal Areas (90, 50, 25 and 10 km2).*

Period of infection	Variable	Hexagon 90 km2			Hexagon 50 km2			Hexagon 25 km2			Hexagon 10 km2		
		P-value	RR	95% CI	P-value	RR	95% CI	P-value	RR	95% CI	P-value	RR	95% CI
Winter	Gender												
	Males					*Reference category*							
	Females	0.20	1.30	0.87-1.94	0.20	1.30	0.87-1.94	0.20	1.30	0.87-1.94	0.20	1.30	0.87-1.93
	Age category (years)												
	0-4	**<0.01**	**2.80**	**1.39-5.35**	**<0.01**	**2.80**	**1.39-5.35**	**<0.01**	**2.81**	**1.39-5.36**	**<0.01**	**2.80**	**1.39-5.36**
	5-9	**<0.01**	**2.82**	**1.43-5.30**	**<0.01**	**2.82**	**1.43-5.31**	**<0.01**	**2.82**	**1.43-5.30**	**<0.01**	**2.82**	**1.43-5.31**
	10-49	0.67	1.11	0.70-1.78	0.67	1.11	0.70-1.78	0.66	1.11	0.70-1.79	0.66	1.11	0.70-1.78
	≥ 50					*Reference category*							
	Type of animal [a]												
	Small ruminants	0.11	1.15	0.97-1.37	0.14	1.12	0.96-1.30	0.07	1.12	0.99-1.27	0.58	0.97	0.88-1.08
	Cattle	0.39	0.92	0.75-1.12	0.34	0.93	0.79-1.08	0.66	0.97	0.86-1.10	0.19	1.07	0.97-1.19
	Poultry	0.91	1.00	0.94-1.07	0.93	1.00	0.94-1.06	0.56	1.01	0.97-1.07	0.89	1.00	0.95-1.04
	Pigs	0.51	1.03	0.94-1.13	0.27	1.05	0.96-1.14	0.79	0.99	0.92-1.07	0.65	0.99	0.92-1.05

[a] *Population-weighted number of animals*

13

317   Living in an hexagonal area of 90 km2 with twice as much population-weighted small
318   ruminants increased the incidence rate of reporting STEC O157 infection in summer, with a RR
319   of 1.13 (95% CI 1.01-1.26) (Table 2). Other hexagonal areas have comparable results, except the
320   one of 10 km2. Here, small ruminants were not significantly associated with STEC O157
321   infections. To further explore this, the analyses at this spatial scale was repeated with goats and
322   sheep separately. The results showed that goats are still significant in summer, with a RR of 1.07
323   (95% CI 1.01-1.3), while sheep no longer pose a risk. In both analyses, pigs are marginally
324   associated with STEC O157 infections, with similar RRs. As other studies showed a clear
325   association with cattle density per municipality in summer, the analyses were repeated with only
326   cattle for hexagonal areas of 90 km2. Here, the population-weighted number of cattle only had a
327   marginal significant association with human STEC O157, with a RR of 1.08 (95% CI 1.00-1.17).
328   In winter, none of the animal types were associated with STEC O157 infections (Table 3).
329   Poultry was never associated with STEC O157 infection. As the population-weighted number of
330   small ruminants in an area was the only consistent significant risk factor for different spatial
331   scales within this study, the PAF was calculated for this factor only. The population-weighted
332   number of small ruminants had a PAF of 49% (95% CI of 8%-72%).
333   The variation in the spatially structured residual risks of the main model showed some
334   dependence on region and period of infection (Figure 4), with a slightly increased residual risk
335   for STEC O157 infection in the northern, mid-eastern and south-western regions of the
336   Netherlands in winter and in the mid-eastern region in summer. A lower residual risk was found
337   in the mid-west to north-west and the south-east region for both periods of infection.
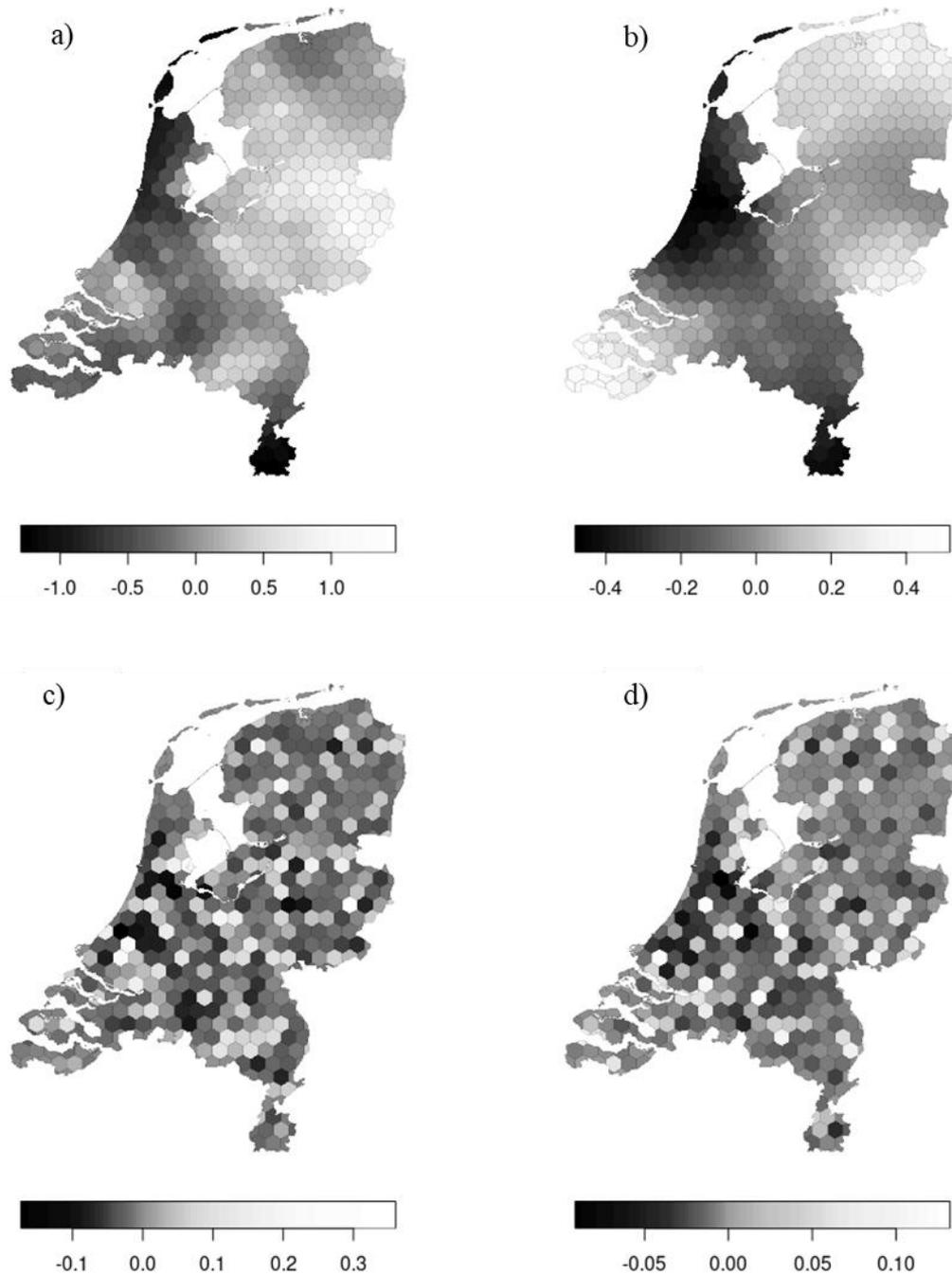
338
Figure 4.

340 Maps of the spatially structured variation modelled by the Conditional Autoregressive Model
341 (CAR) in summer (a) and winter (b) and maps of the spatially unstructured variation modelled
342 by independent and identically distributed (IID) Gaussian noise in summer (a) and winter (b) for
343 hexagonal areas of 90 km2.

5 Discussion

The aim of this study was to explore the spatial association between sporadic human STEC O157 infections and the exposure to livestock (small ruminants, cattle, poultry, and pigs) in the Netherlands, a country with high densities of humans and livestock animals, for the years 2007-2016. This was done using a state-of-the-art spatial analysis method, in which hexagonal areas were used in combination with a novel metric that was applied to define the probability of exposure: the population-weighted number of animals per hexagon.

Results showed that there is a consistent significant association between the population-weighted number of small ruminants and the incidence of reported human STEC O157 infections in summer with a PAF of 49%. This means that in the absence of environmental exposure to small ruminants, the number of sporadic STEC O157 infections due to environmental transmission should be expected to decrease by 49%, although the uncertainty ranges between 8% and 72%. Since we only accounted for livestock density as a risk factor for infection with O157, we were not able to quantify the relative importance of environmental transmission within the broader context of all possible transmission routes, such as food. The risk of ecological biases could also not be quantified in this regard. Although we corrected for age and gender at an individual level, areas can still differ regarding confounders that are not included in our analyses, as is also suggested by the maps of the spatially structured variation (CAR).

The finding that small ruminants are important contributors to human STEC O157 infections is supported by a Dutch source attribution study (Mughini-Gras et al., 2018), which shows that while domestic ruminants (cattle, sheep, and goats) are responsible for approximately three-quarters of reported human STEC (all serotypes) infections, small ruminants in particular accounted for 25% of all STEC infections. In the Netherlands, STEC O157 has been isolated from sheep and goats (Heuvelink et al., 1998; Heuvelink et al., 2002). Additionally, STEC was detected at almost all dairy goat and sheep farms in the Netherlands that were included in the Dutch surveillance of zoonoses in 2016, although STEC O157 was only detected at one farm (Opsteegh 2018). This reflects a common paradox regarding the results of animal sampling, in which small ruminants are generally considered as a primary reservoir for STEC O157, but their occurrence is infrequently demonstrated. This may reflect the sporadic and/or intermittent nature of STEC O157 carriage and low numbers of bacteria residing in colonized animals, or insufficiently sensitive sampling and culturing approaches (Ferens & Hovde, 2011).

Given the presence of STEC in small ruminants' feces and farms, it is plausible that human infections occur via environmental transmission. In the Netherlands, small ruminants are usually kept in deep litter houses, with partially open walls or roofs (Schimmer et al., 2011). An initial layer of litter (usually straw or sawdust material) is spread for the animals to use for bedding material and to defecate on. As soon as this layer is soiled, new layers are added, which can build up to a depth of 1-2 meters. This process generates a lot of dust, which is easily spread into the environment through the often (partially) open housing system. As a result, the transport of STEC O157 in dust through the air can be one of the possible environmental transmission routes if infected animals are present on the farm (Chase-Topping et al., 2008; Schimmer et al., 2011). The plausibility of air-borne transmission is supported by a study focusing on microbial air pollution from livestock farms in the Netherlands, where a higher concentration of commensals, amongst which *Escherichia coli*, in dust particles was found in rural areas with higher farm density (de Rooij et al., 2019). Although no significant associations with the number of goats and sheep were found, the presence of livestock-related microbial markers, such as *Escherichia coli*, indicates that microbial air pollution with *Escherichia coli* is reasonable. The same phenomenon

390   was observed for Campylobacter, which coincides with a higher Campylobacter incidence in
391   poultry-dense areas, the main reservoir of Campylobacter (de Rooij et al., 2019; Poulsen et al.,
392   2018). Furthermore, transmission of STEC O157 to humans may occur via soil or water, since
393   dust precipitates and the stable litter that is stored outside the stable comes into contact with soil
394   and possibly fresh water systems through washout after heavy rainfall (Elson et al., 2018).
395   Whilst several studies, including a Dutch one, showed a significant spatial association
396   between cattle and STEC O157 infections (Friesema et al., 2011; ÓHaiseadha et al., 2017;
397   Norval J. C. Strachan et al., 2006; Widgren et al., 2018), we did not. This could have several
398   possible explanations.
399   First, a major difference is the inclusion of small ruminants in this study, next to cattle. Cattle
400   farms are widely distributed in the Netherlands, while small ruminants have a more profound
401   environmental spread. To study whether this could lead to different results, the analyses were
402   performed for a model with only cattle. The results showed that the population-weighted number
403   of cattle had a marginal significant association with human STEC O157, while this effect is not
404   significant anymore after the inclusion of pigs, poultry and small ruminants. This might indicate
405   that the spatial association observed for cattle could be due to its spatial relatedness with small
406   ruminants, the latter which may play a more important role in environmental STEC O157
407   transmission. This proves that it is meaningful to look at the combined effects of all possible
408   reservoirs for STEC O157. Such a combined analysis is especially important in a country like the
409   Netherlands, which has a peculiar situation in terms of livestock and population density as
410   compared to other countries (Smit & Heederik, 2017). Indeed, it is one of the most densely
411   populated countries in the world in combination with a high density of intensive livestock farms
412   (Smit & Heederik, 2017). An example of such a situation is the Q-fever epidemic in the
413   Netherlands (Schimmer et al., 2011), which became an epidemic because most goat farms were
414   located very close to locations with a high population density (Schimmer et al., 2011). As all the
415   different types of livestock farms in the Netherlands are intertwined and mixed throughout the
416   landscape, spatial inter-relatedness with other animal species does play a role (de Rooij et al.,
417   2019). This makes it complicated to disentangle the effects and to look at each type of livestock
418   separately, emphasizing that a more complete model in terms of possible reservoirs of STEC
419   O157 is necessary for a proper analysis (de Rooij et al., 2019).
420   Second, livestock farming in the Netherlands underwent several changes in the past few
421   years that could explain the different findings as well (Bos et al., 2013). There was a reduction in
422   the number of farms over the years, which was paralleled by an increase in the number of
423   animals per farm, with cows being increasingly kept inside throughout the year (Bos et al., 2013;
424   Groot & van't Hooft, 2016; Smit & Heederik, 2017). As cattle is more often kept inside and their
425   housing is closed, it is possible that aerial spread of STEC from cattle is reduced over the years
426   and that small ruminants play a more important role nowadays.
427   Third, this study used a different spatial metric as response variable in order to do the spatial
428   regression analyses on the hexagonal grid. Here, the population-weighted number of animals was
429   used instead of animal density as exposure measure to transform the spatial data from one
430   regional division to the other (Elson et al., 2018; Friesema et al., 2011; Hallisey et al., 2017;
431   ÓHaiseadha et al., 2017). However, our approach has the advantage over areal weighted
432   interpolation that it can more accurately estimate the population demographics in transforming
433   small counts by four-digit postal code regions to aggregated counts for large, non-standard study
434   zones (hexagons) (Hallisey et al., 2017). Moreover, because the probability of exposure on a
435   population level is not only determined by the number of animals in a certain area, but also by

436 the number of residents in a certain area and where they live inside an area, this study is more
437 likely to have captured true environmental exposure, as exposure is less likely to occur when
438 nobody lives in the vicinity of these animals (Mulder et al., 2016). Furthermore, in contrast to
439 other studies, we took into account potential exposure to animals in neighboring hexagons,
440 because pathogen spread is not hold back by "invisible" hexagonal boundaries.
441   In this study, no associations were found between poultry, pigs and STEC O157 infections in
442 the multivariable model. This supports the finding that STEC has been isolated only sporadically
443 from animals other than ruminants and these animals can merely be seen as spill-over hosts
444 (Caprioli et al., 2005; Mughini-Gras et al., 2018). Also, a low estimated contribution to human
445 STEC infections has previously been found for poultry and pigs in the Netherlands (Mughini-
446 Gras et al., 2018). However, pigs did show a positive association with human STEC infections at
447 a hexagonal size of 10 km2 in the multivariable model. This could be due to several reasons,
448 such as limitations of power and more limited exposure metric contrasts at this smaller spatial
449 scale (de Rooij et al., 2019).
450   The association between small ruminants and human STEC O157 infections was only present
451 in the summer. This is in agreement with the incidence of human STEC O157 infections being
452 highest in summer, as well as the seasonality of fecal excretion of STEC in farm animals
453 (Friesema et al., 2011; Heuvelink et al., 1998). Furthermore, humans are more likely to have
454 direct or indirect contact with animals in summer as they probably spend more time outside
455 (Friesema et al., 2011). Similar to what is described globally, women had a higher risk than men
456 to acquire a STEC O157 infection in summer and the incidence of STEC O157 was highest in
457 children <10 years and strongest in children <5 years (Elson et al., 2018; Friesema et al., 2011).
458   Compared to a previous Dutch study (Friesema et al., 2011), underreporting of the human
459 STEC O157 infections and the geographical laboratory bias did not change. Human STEC O157
460 cases included in this study likely represent the more severe cases, as mild cases often go
461 unnoticed, because they may not always seek medical attention or do not get laboratory tested
462 and hence, do not end up in the surveillance records (Friesema et al., 2011; van den Brandhof et
463 al., 2006). The laboratory surveillance is based on a voluntary system, but despite the fact that
464 the notification is mandatory, it is not guaranteed that all laboratories send in their isolates on a
465 regular basis.

466 **6 Conclusions**
467   Results of this study indicate that living in proximity of small ruminants, is a risk factor for
468 acquiring STEC O157 infection, possibly due to environmental transmission via the air by
469 microbial contamination of dust particles with STEC O157, soil or water. However, the exact
470 underlying mechanisms warrant further investigation, and could offer new targets for control.
471 The finding that small ruminants, and not cattle, are significantly associated with human STEC
472 O157 infection is in contradiction with earlier studies. It could be explained by the inclusion of
473 small ruminants in the analysis, a changing farming landscape over the years, and the newly
474 developed exposure metric, the population-weighted number of animals per hexagon, which
475 showed potential to improve existing spatial modelling studies on infectious diseases related to
476 livestock exposure, especially in densely populated regions.
477

478 **Abbreviations**

479 **STEC O157** = Shiga Toxin-producing Escherichia coli
480 **PAF** = Population attributable fraction
481 **INLA** = Integrated nested Laplace approximation
482 **EU** = European Union
483 **CAR** = Conditional Autoregressive Model
484 **IID** = Independent and identically distributed
485 **RR** = Rate ratio
486 **MHS** = Municipal Health Service
487 **RIVM** = Dutch National Institute for Public Health and the Environment
488 **PCR** = Polymerase chain reaction
489 **Stx$_1$** = Shiga Toxin 1
490 **Stx$_2$** = Shiga Toxin 2
491 **CI** = Confidence interval

492 **Acknowledgments**

493   The authors would like to thank Ingrid Friesema for her valuable input about the STEC
494 O157 surveillance in the Netherlands.

495 **Funding**

496   Not applicable

497 **Conflict of interest**

498   None declared

499 **Data availability**

500   Links and citations to most of the data used in this study can be found in the main text.
501 The four-digit postal code region shapefiles and the six-digit postal code point locations of the
502 Netherlands that were used within this study were obtained by the RIVM from the company: Iris
503 International. Those shapefiles can only be given to those for whom permission has been granted
504 by this company. They can be reached at this address: Gr.v. Prinstererlaan 20, 2271 EN,
505 Voorburg, the Netherlands. Tel: +31(0)70-3863891, fax: +31(0)70-3873625, e-mail: info@iris-
506 int.nl .

507   The STEC O157 case data are available within OSIRIS, the Dutch surveillance system
508 and only researchers within the RIVM with access to this database can use those data as it
509 contains privacy sensitive information of cases and therefore are not accessible to the public or
510 research community following the legislation of the Dutch law and the Dutch Data Protection
511 Authority (Dutch Data Protection Authority, 2020a, 2020b).

512 **Authors' contributions**

513   Conceptualization – ACM, JvdK, RP, LMG, EF; Data curation – ACM, JvdK; Formal
514 analysis: ACM, Funding acquisition: EF, LMG; Investigation – ACM, JvdK, DH, RP, LMG, EF;
515 Methodology – ACM, JvdK; Project administration – ACM, EF; Resources – JvdK, EF, LMG;

516   Software – JvdK, ACM; Supervision – EF, LMG, DH; Validation – JvdK; Visualization – ACM;
517   Writing – original draft – ACM; Writing – review & editing – ACM, JvdK, DH, RP, LMG, EF.

References

519   Arsenault, J., Michel, P., Berke, O., Ravel, A., & Gosselin, P. (2013). How to choose geographical units in
520   ecological studies: proposal and application to campylobacteriosis. *Spatial and spatio-temporal*
521   *epidemiology, 7*, 11-24. doi:https://doi.org/10.1016/j.sste.2013.04.004
522   Arya, S., Mount, D., Kemp, S. E., & Jefferis, G. (2015). RANN: Fast Nearest Neighbour Search (Wraps Arya and
523   Mount's ANN Library). Retrieved from http://cran.rapporter.net/bin/macosx/contrib/3.2/PACKAGES
524   Bates, D., Maechler, M., Davis, T. A., Oehlschlägel, J., & Riedy, J. (2019). Matrix: Sparse and Dense Matrix
525   Classes and Methods. Retrieved from https://cran.r-project.org/web/packages/Matrix/index.html
526   Berry, E. D., Wells, J. E., Bono, J. L., Woodbury, B. L., Kalchayanand, N., Norman, K. N., et al. (2015). Effect of
527   Proximity to a Cattle Feedlot on O157:H7 Contamination of Leafy Greens and Evaluation of the Potential
528   for Airborne Transmission. *Applied and Environmental Microbiology, 81*(3), 1101-1110.
529   doi:https://doi.org/10.1128/AEM.02998-14
530   Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics.
531   *Annals of the institute of statistical mathematics, 43*(1), 1-20. doi:https://doi.org/10.1007/BF00116466
532   Birch, C. P. D., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation,
533   experiment and simulation in ecology. *Ecological Modelling, 206*(3), 347-359.
534   doi:https://doi.org/10.1016/j.ecolmodel.2007.03.041
535   Birch, C. P. D., Vuichard, N., & Werkman, B. R. (2000). Modelling the effects of patch size on vegetation
536   dynamics: Bracken [Pteridium aquilinum (L.) Kuhn] under grazing. *Annals of Botany, 85*, 63-76.
537   doi:https://doi.org/10.1006/anbo.1999.1081
538   Bivand, R., Altman, M., Anselin, L., Assunção, R., Berke, O., Bernat, A., et al. (2019). spdep: Spatial Dependence:
539   Weighting Schemes, Statistics. Retrieved from https://cran.r-project.org/web/packages/spdep/index.html
540   Bos, J. F., Smit, A. B. L., & Schröder, J. J. (2013). Is agricultural intensification in The Netherlands running up to its
541   limits? *NJAS-Wageningen Journal of Life Sciences, 66*, 65-73.
542   doi:https://doi.org/10.1016/j.njas.2013.06.001
543   Caprioli, A., Morabito, S., Brugère, H., & Oswald, E. (2005). Enterohaemorrhagic Escherichia coli: emerging issues
544   on virulence and modes of transmission. *Veterinary research, 36*(3), 289-311.
545   doi:https://doi.org/10.1051/vetres:2005002
546   CBS. (2019a). Bevolking; geslacht, leeftijd en viercijferige postcode, 1 januari. Retrieved from
547   https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83502NED/table?ts=1575556895525
548   CBS. (2019b). Landbouwtelling. Retrieved from https://www.cbs.nl/nl-nl/onze-
549   diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/landbouwtelling
550   Chase-Topping, M., Gally, D., Low, C., Matthews, L., & Woolhouse, M. (2008). Super-shedding and the link
551   between human infection and livestock carriage of Escherichia coli O157. *Nature Reviews Microbiology,*
552   *6*(12), 904. doi:https://doi.org/10.1038/nrmicro2029
553   De Jonge, E., & Houweling, S. (2019). cbsodataR: Statistics Netherlands (CBS) Open Data API Client. Retrieved
554   from https://rdrr.io/cran/cbsodataR/
555   de Rooij, M. M. T., Hoek, G., Schmitt, H., Janse, I., Swart, A., Maassen, C. B. M., et al. (2019). Insights into
556   Livestock-Related Microbial Concentrations in Air at Residential Level in a Livestock Dense Area.
557   *Environmental Science & Technology, 53*(13), 7746-7758. doi:https://doi.org/10.1021/acs.est.8b07029
558   Dutch Data Protection Authority. (2020a). Retrieved from https://www.autoriteitpersoonsgegevens.nl/en
559   Dutch Data Protection Authority. (2020b). Retrieved from https://www.autoriteitpersoonsgegevens.nl/en/about-
560   dutch-dpa/tasks-and-powers-dutch-dpa
561   Elson, R., Grace, K., Vivancos, R., Jenkins, C., Adak, G. K., O'Brien, S. J., & Lake, I. R. (2018). A spatial and
562   temporal analysis of risk factors associated with sporadic Shiga toxin-producing Escherichia coli O157
563   infection in England between 2009 and 2015. *Epidemiol Infect, 146*(15), 1928-1939.
564   doi:https://doi.org/10.1017/S095026881800256X
565   European Centre for Disease Prevention and Control. (2019). *Annual Epidemiological report for 2017 - Shiga*
566   *toxin/verocytotoxin-producing Escherichia coli (STEC/VTEC) infection - Annual Epidemiological Report*
567   *for 2017*. Retrieved from Stockholm: https://www.ecdc.europa.eu/en/publications-data/shiga-
568   toxinverocytotoxin-producing-escherichia-coli-stecvtec-infection-annual

569 European Commission - Eurostat. (2019). Nuts - Nnuts - Nomenclature of territorial units for statistics. Retrieved
570   from https://ec.europa.eu/eurostat/web/nuts/background
571 European Food Safety Authority, & European Centre for Disease Prevention and Control. (2019). The European
572   Union One Health 2018 Zoonoses Report. *EFSA Journal, 17*(12), e05926.
573   doi:https://doi.org/10.2903/j.efsa.2019.5926
574 Ferens, W. A., & Hovde, C. J. (2011). Escherichia coli O157: H7: animal reservoir and sources of human infection.
575   *Foodborne pathogens and disease, 8*(4), 465-487. doi:https://doi.org/10.1089/fpd.2010.0673
576 Franz, E., Rotariu, O., Lopes, B. S., MacRae, M., Bono, J. L., Laing, C., et al. (2018). Phylogeographic analysis
577   reveals multiple international transmission events have driven the global emergence of Escherichia coli
578   O157: H7. *Clinical Infectious Diseases, 69*(3), 428-437. doi:https://doi.org/10.1093/cid/ciy919
579 Franz, E., Schijven, J., de Roda Husman, A. M., & Blaak, H. (2014). Meta-Regression Analysis of Commensal and
580   Pathogenic Escherichia coli Survival in Soil and Water. *Environmental Science & Technology, 48*(12),
581   6763-6771. doi:https://doi.org/10.1021/es501677c
582 Friesema, I., Kuiling, S., van der Voort, M., in 't Veld, P. H., Heck, M. E. O. C., & Franz, E. (2017). Surveillance
583   van shigatoxineproducerende Escherichia coli (STEC) in Nederland, 2016. Retrieved from
584   https://www.rivm.nl/surveillance-van-shigatoxineproducerende-escherichia-coli-stec-in-nederland-2016
585 Friesema, I., Sigmundsdottir, G., Van Der Zwaluw, K., Heuvelink, A., Schimmer, B., De Jager, C., et al. (2008). An
586   international outbreak of Shiga toxin-producing Escherichia coli O157 infection due to lettuce, September–
587   October 2007. *Euro Surveill., 13*(50), 19065. doi:https://doi.org/10.2807/ese.13.50.19065-en
588 Friesema, I., Van De Kassteele, J., De Jager, C. M., Heuvelink, A. E., & Van Pelt, W. (2011). Geographical
589   association between livestock density and human Shiga toxin-producing Escherichia coli O157 infections.
590   *Epidemiol. Infect., 139*(7), 1081-1087. doi:https://doi.org/10.1017/S0950268810002050
591 Greenland, K., De Jager, C., Heuvelink, A., Van Der Zwaluw, K., Heck, M., Notermans, D., et al. (2009).
592   Nationwide outbreak of STEC O157 infection in the Netherlands, December 2008-January 2009:
593   continuous risk of consuming raw beef products. *Euro Surveill., 14*(8). Retrieved from
594   http://researchonline.lshtm.ac.uk/id/eprint/748775/
595 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software,
596   40*(3), 1-25.
597 Groot, M. J., & van't Hooft, K. E. (2016). The Hidden Effects of Dairy Farming on Public and Environmental
598   Health in the Netherlands, India, Ethiopia, and Uganda, Considering the Use of Antibiotics and Other
599   Agro-chemicals. *Frontiers in Public Health, 4*(12). doi:https://doi.org/10.3389/fpubh.2016.00012
600 Hallisey, E., Tai, E., Berens, A., Wilt, G., Peipins, L., Lewis, B., et al. (2017). Transforming geographic scale: a
601   comparison of combined population and areal weighting to other interpolation methods. *International
602   Journal of Health Geographics, 16*(1), 29. doi:https://doi.org/10.1186/s12942-017-0102-z
603 Heuvelink, A., Van Den Biggelaar, F., De Boer, E., Herbes, R., Melchers, W., Huis, J., & Monnens, L. (1998).
604   Isolation and characterization of verocytotoxin-producing Escherichia coli O157 strains from Dutch cattle
605   and sheep. *Journal of Clinical Microbiology, 36*(4), 878-882. doi:https://doi.org/10.1128/JCM.36.4.878-
606   882
607 Heuvelink, A., Van Heerwaarden, C., Zwartkruis-Nahuis, J., Van Oosterom, R., Edink, K., Van Duynhoven, Y., &
608   De Boer, E. (2002). Escherichia coli O157 infection associated with a petting zoo. *Epidemiology &
609   Infection, 129*(2), 295-302. doi:https://doi.org/10.1017/S095026880200732X
610 Keitt, T. H. (2010). rgdal: Bindings for the Geospatial Data Abstraction Library, R package version 0.6-28.
611   *http://cran. r-project. org/package= rgdal.*
612 Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*: Chapman and
613   Hall/CRC.
614 Mughini-Gras, L., van Pelt, W., van der Voort, M., Heck, M., Friesema, I., & Franz, E. (2018). Attribution of human
615   infections with Shiga toxin-producing Escherichia coli (STEC) to livestock sources and identification of
616   source-specific risk factors, The Netherlands (2010–2014). *Zoonoses and Public Health, 65*(1), e8-e22.
617   doi:https://doi.org/10.1111/zph.12403
618 Mulder, A. C., Snabilie, M., & Braks, M. A. (2016). From guessing to GIS-ing: empowering land managers. In
619   *Ecology and prevention of Lyme borreliosis* (Vol. 4, pp. 373-387): Wageningen Academic Publishers.
620 Neuwirth, E. (2015). Package 'RColorBrewer'. Retrieved from https://cran.r-
621   project.org/web/packages/RColorBrewer/RColorBrewer.pdf
622 ÓHaiseadha, C., Hynds, P. D., Fallon, U. B., & O'Dwyer, J. (2017). A geostatistical investigation of agricultural and
623   infrastructural risk factors associated with primary verotoxigenic E. coli (VTEC) infection in the Republic

624   of Ireland, 2008–2013. *Epidemiology and Infection, 145*(1), 95-105.
625   doi:https://doi.org/10.1017/S095026881600193X
626 Pebesma, E. (2019). st_make_grid. Retrieved from https://www.rdocumentation.org/packages/sf/versions/0.8-
627   0/topics/st_make_grid
628 Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., et al. (2019). sf: Simple Features for R.
629   Retrieved from https://cran.r-project.org/web/packages/sf/index.html
630 Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R., Sumner, M., et al. (2019). sp: Classes and
631   Methods for Spatial Data. Retrieved from https://cran.r-project.org/web/packages/sp/index.html
632 Poulsen, M. N., Pollak, J., Sills, D. L., Casey, J. A., Rasmussen, S. G., Nachman, K. E., et al. (2018). Residential
633   proximity to high-density poultry operations associated with campylobacteriosis and infectious diarrhea.
634   *International Journal of Hygiene and Environmental Health, 221*(2), 323-333.
635   doi:https://doi.org/10.1016/j.ijheh.2017.12.005
636 R-Core. (2017). parallel v3.6.0. Retrieved from https://www.rdocumentation.org/packages/parallel/versions/3.6.0
637 RCT. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing.
638   Vienna, Austria. https://www.R-project.org
639 Rue, H. (2019). The R-INLA project. Retrieved from http://www.r-inla.org/home
640 Rue, H., Lindgren, F., Simpson, D., Martino, S., Teixeira Krainski, E., Bakka, H., et al. (2019). INLA: Full Bayesian
641   Analysis of Latent Gaussian Models using Integrated Nested Laplace Approximations. *R package version,*
642   *19*(03).
643 RVO. (2019). Gecombineerde opgave. Retrieved from https://mijn.rvo.nl/gecombineerde-opgave
644 Schimmer, B., Luttikholt, S., Hautvast, J. L., Graat, E. A., Vellema, P., & van Duynhoven, Y. T. (2011).
645   Seroprevalence and risk factors of Q fever in goats on commercial dairy goat farms in the Netherlands,
646   2009-2010. *BMC Veterinary Research, 7*(1), 81. doi:https://doi.org/10.1186/1746-6148-7-81
647 Shafran-Nathan, R., Levy, I., Levin, N., & Broday, D. M. (2017). Ecological bias in environmental health studies:
648   the problem of aggregation of multiple data sources. *Air Quality, Atmosphere & Health, 10*(4), 411-420.
649   doi:https://doi.org/10.1007/s11869-016-0436-x
650 Smit, L. A. M., & Heederik, D. (2017). Impacts of Intensive Livestock Production on Human Health in Densely
651   Populated Regions. *GeoHealth, 1*(7), 272-277. doi:https://doi.org/10.1002/2017GH000103
652 Strachan, N. J. C., Dunn, G. M., Locking, M. E., Reid, T. M. S., & Ogden, I. D. (2006). Escherichia coli O157:
653   Burger bug or environmental pathogen? *International Journal of Food Microbiology, 112*(2), 129-137.
654   doi:https://doi.org/10.1016/j.ijfoodmicro.2006.06.021
655 Strachan, N. J. C., Fenlon, D. R., & Ogden, I. D. (2001). Modelling the vector pathway and infection of humans in
656   an environmental outbreak of Escherichia coli O157. *FEMS Microbiology Letters, 203*(1), 69-73.
657   doi:https://doi.org/10.1111/j.1574-6968.2001.tb10822.x
658 van den Brandhof, W. E., Bartelds, A. I., Koopmans, M. P., & van Duynhoven, Y. T. (2006). General practitioner
659   practices in requesting laboratory tests for patients with gastroenteritis in the Netherlands, 2001–2002.
660   *BMC Family Practice, 7*(1), 56. doi:https://doi.org/10.1186/1471-2296-7-56
661 Wickham, H. (2019). stringr v1.4.0. Retrieved from https://www.rdocumentation.org/packages/stringr/versions/1.4.0
662 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the
663   Tidyverse. *Journal of Open Source Software, 4*(43), 1686.
664 Wickham, H., Bryan, J., Rstudio, Kalicinski, M., Valery, K., Leitienne, C., et al. (2019). Package 'readxl'. Retrieved
665   from https://cran.r-project.org/web/packages/readxl/readxl.pdf
666 Wickham, H., Francois, R., Henry, L., & Müller, K. (2019). Package 'dplyr'. Retrieved from https://cran.r-
667   project.org/web/packages/dplyr/dplyr.pdf
668 Wickham, H., Henry, L., & Rstudio. (2019). tidyr: Tidy Messy Data. Retrieved from https://cran.r-
669   project.org/web/packages/tidyr/index.html
670 Widgren, S., Engblom, S., Emanuelson, U., & Lindberg, A. (2018). Spatio-temporal modelling of verotoxigenic
671   Escherichia coli O157 in cattle in Sweden: exploring options for control. *Veterinary research, 49*(1), 78.
672   doi:https://doi.org/10.1186/s13567-018-0574-2
673

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:

$$\mu = e^{\beta_0 + \beta_1(x+1)}$$
$$= e^{\beta_0} e^{\beta_1(x+1)}$$

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.

4

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.
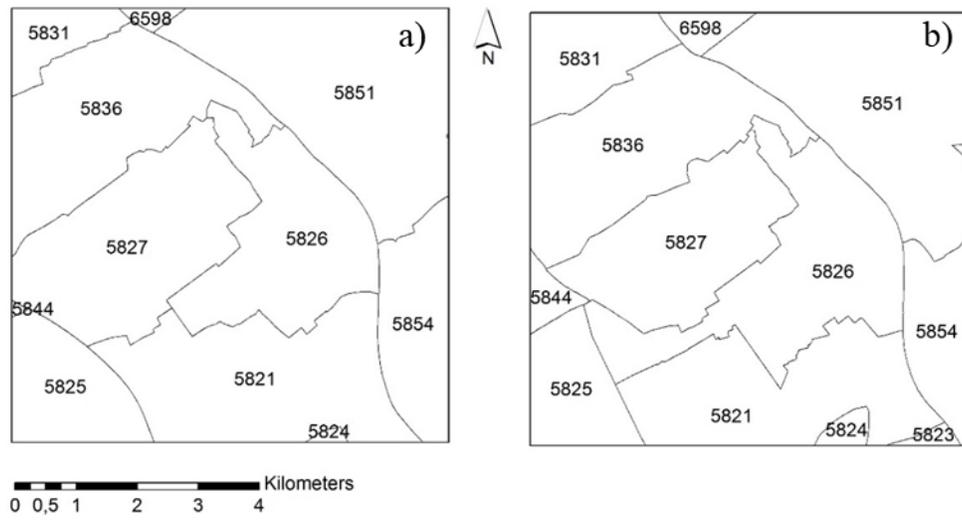
Figure S1. Example of changing four-digit postal code regions of the Netherlands over the years; a) 2009 compared to b) 2016.
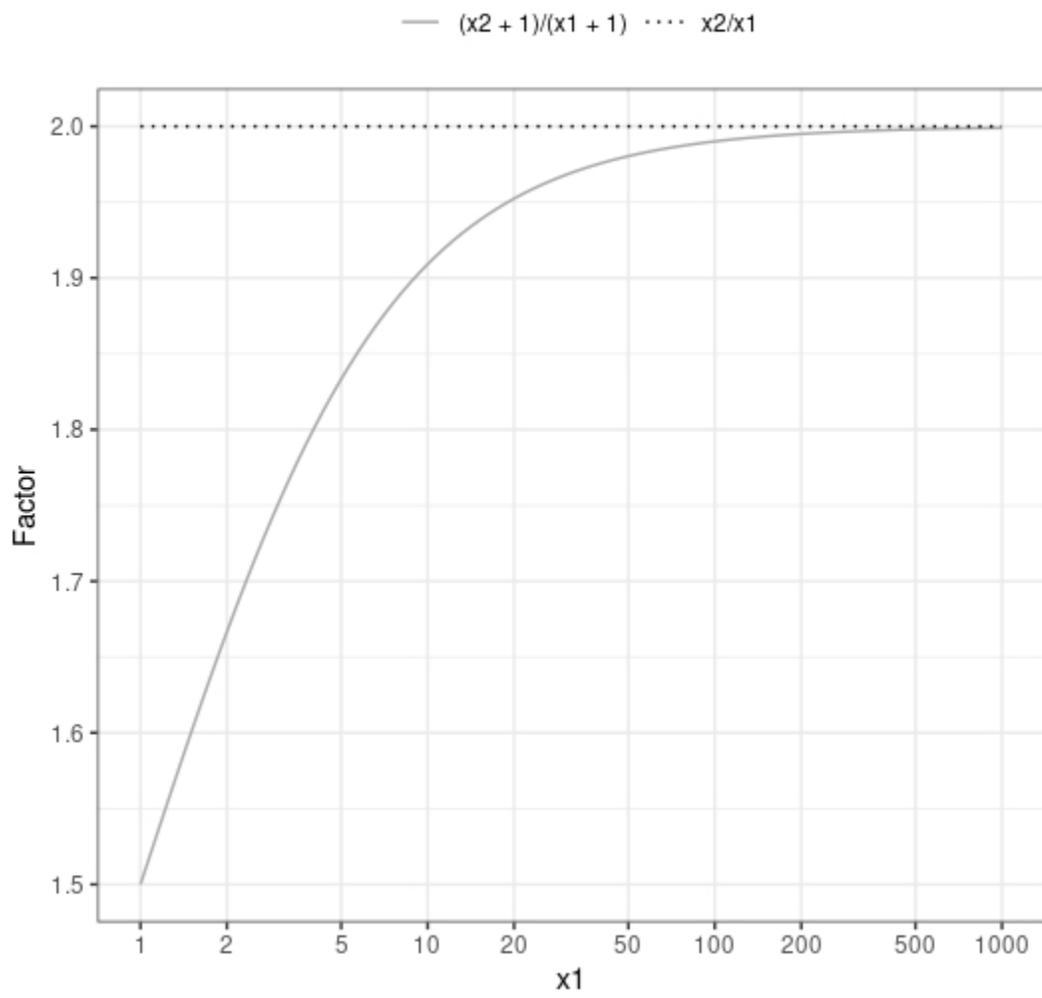
Figure S2. Visualization of interpretation rate ratio (RR) for an exposure measure x_1.
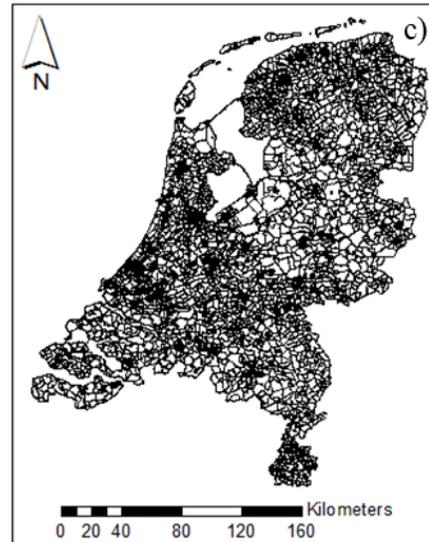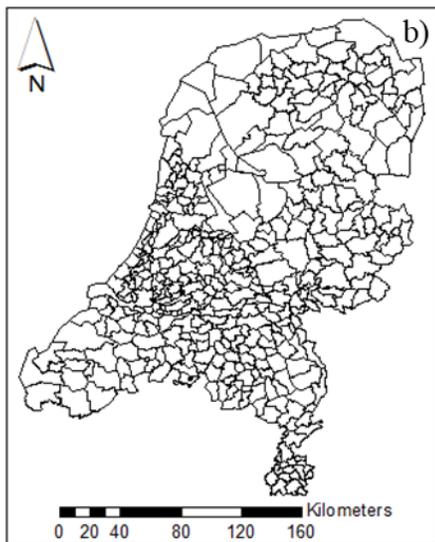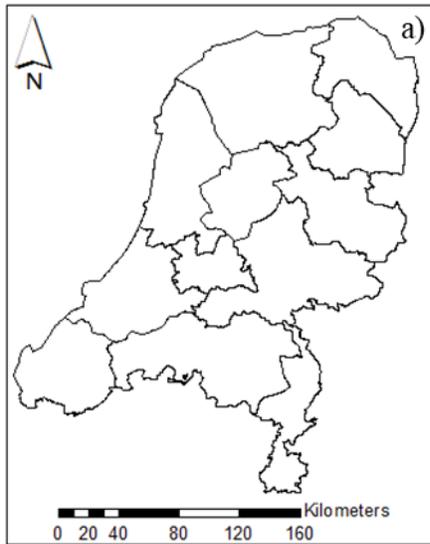
Figure S3. The different administrative boundaries and spatial scales of the Netherlands. a) Provinces (NUTS 2 regions), b) Municipalities, c) Four-digit postal code regions.

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.

Variable	Hexagon 90 km2			Hexagon 50 km2			Hexagon 25 km2			Hexagon 10 km2		
	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).