



Spatio-temporal outbreaks of campylobacteriosis and the role of fresh-milk vending machines in the Czech Republic: A methodological study

Lukáš Marek,¹⁻³ Vít Pászto³

¹Department of Geography, University of Canterbury, Christchurch; ²Geospatial Research Institute, University of Canterbury, Christchurch, New Zealand; ³Department of Geoinformatics, Palacky University, Olomouc, Czech Republic

Abstract

Inspired by local outbreaks of campylobacteriosis in the Czech Republic in 2010 linked to the debate about alleged health risks of the raw milk consumption, a detailed study was carried out. Firstly, scanning was utilised to identify spatio-temporal clusters of the disease from 2008 to 2012. Then a spatial method (geographical profiling originally developed for criminology) served as assessment in selecting fresh-milk vending machines that could have contributed to some of the local campylobacteriosis outbreaks. Even though an area of increased relative risk of the disease was identified in the affected city of České Budějovice during January and February 2010, geoprofiling did not identify any vending machines in the area as the potential source. However, possible sources in some nearby cities were suggested. Overall, 14 high-rate clusters including the localisation of 9% of the vending machines installed in the Czech Republic were found in the period

2008-2012. Although the vending machines are subject to strict hygiene standards and regular testing, a potential link between a small number of them and the spatial distribution of campylobacteriosis has been detected in the Czech Republic. This should be taken into account in public health research of the disease.

Introduction

Fresh-milk vending machines are increasingly popular in the Czech Republic as they offer benefits both for consumers and farmers. Producers have a chance to provide people with fresh farm milk directly, while they might profit financially and even morally. However, the situation changed due to a discussion taking place in the region of South Bohemian in early 2010. During that time, the growing number of fresh-milk vending machines became challenged by reports of about possible health issues (mainly digestive problems) caused by the drinking of fresh milk sold through these vending machines. The originally very positive adoption of fresh-milk vending machines was negatively affected by the announcement by the then Czech chief hygiene officer pointing out to these health risks officially issued by the Ministry of Health in the Czech Republic (MoH CR) (2010). Although legislation and control mechanisms assure monitoring of both providers and machines by the relevant authorities, such as the Agrarian Chamber of the Czech Republic (2010; Hygiene, 2010), the announcement slowed down the expansion of fresh-milk vending machines, even phased out some machines (Andřlová, 2011).

Campylobacter is responsible for infectious diseases that affect humans and animals. *Campylobacter*, typically *C. jejuni* or *C. coli*, are Gram-negative, microaerophilic, curved or spiral rods in the family of Campylobacteriaceae that are ubiquitous and potentially found in humans and many animal species, although some clonal complexes are more commonly associated with certain hosts (Center for Food Security & Public Health, 2013). Although campylobacteriosis is the most common human gastroenteritis in developed countries (Weisent *et al.*, 2011), this infection is not as well-known by the general public as many others, *e.g.*, salmonellosis, as noted by the Institute of Health Information and Statistics of the Czech Republic (2013). The agent is usually transmitted by ingestion of contaminated poultry meat, but according to the Centers for Disease Control and Prevention (CDC), outbreaks can also be linked to raw milk and dairy products (Centers for Disease Control and Prevention, 2009; LeJeune and Rajala-Schultz, 2009), contaminated water, or by exposure to infected animals (livestock, pets) and can rarely also be caused by interpersonal transmission (Ambrožová, 2011), while up to 75% of chicken meat on the market can be contami-

Correspondence: Lukáš Marek, Department of Geography, University of Canterbury, Private Bag 4800, 8140 Christchurch, New Zealand.
E-mail: lukas.marek@canterbury.ac.nz

Key words: Geoprofiling; Spatiotemporal clustering; Fresh milk; Campylobacteriosis; Czech Republic.

Acknowledgements: the authors gratefully acknowledge the support by the Operational Program Education for Competitiveness - European Social Fund (project CZ.1.07/ 2.3.00/20.0170 of the Ministry of Education, Youth and Sports of the Czech Republic) and also by the ERASMUS+ project no. 2016-1-CZ01-KA203-024040 (Spationomy) funded by the European Union. We also thank the National Institute of Public Health for providing the data for the study. The authors are appreciative of reviewers' constructive feedback, and also thank editors for improving the readability of the manuscript.

Received for publication: 5 April 2017.

Revision received: 21 July 2017.

Accepted for publication: 28 July 2017.

©Copyright L. Marek and V. Pászto, 2017

Licensee PAGEPress, Italy

Geospatial Health 2017; 12:572

doi:10.4081/gh.2017.572

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

nated by *Campylobacter* spp. and chilled chickens represent a greater risk to the consumer than frozen chickens, according to veterinary data (Ambrožová, 2011). Although a large number of transmission routes has been identified, the epidemiology of the disease is still unclear, and only half of all cases has been sufficiently described with the infectious route identified (Nylen *et al.*, 2002; Ekdahl *et al.*, 2005). The source of no more than 43% of *Campylobacter* infections has been identified in the Czech Republic during the study period (Figure 1B).

Infected people suffer diarrhoea, fatigue, fever, abdominal pain and nausea, including vomiting. The incubation period is usually 2-5 days (range 1-10 days), while symptoms last one week on average. The demographic groups most vulnerable to campylobacteriosis are children, teenagers and young adults up to 30 years of age (independently of gender) (Lexová *et al.*, 2013) (Figure 1A). The incidence in the Czech Republic reached 230 cases per 100,000 people in 2013 comparing to the incidence of salmonellosis that has stabilised around 100 cases per 100,000 since 2008 (Institute of Health Information and Statistics of the Czech Republic, 2013). Outbreaks in the Czech Republic follow a seasonal pattern peaking during the summer months (mainly August) (Lexová *et al.*, 2013; Domasová, 2014). The infection distribution in an affected population is shown in the series of graphs in Figure 1. Even though the campylobacteriosis is considered the most common reason for gastroenteritis, it is also highly underreported due to both occasional mild symptoms and insufficient diagnostics tests. The study of Havelaar *et al.* (2013) estimates that this infection might be up to 11.3 times more common than reported in the Czech Republic and up to 46.7 times more common in the European Union as a whole.

Mullner *et al.* (2010) applied an individual model to assess the spatial risk of infection based on demography, patient genotype and the potential source of the outbreak. The direct impact of the climate on the geographical distribution of the disease has been studied by Sari Kovats *et al.* (2005) and Jagai *et al.* (2007). Modelling of the relative risk (RR) or incidence of campylobacteriosis, and identification of environmental, demographic or socio-

economic associations has been the main topic of studies in Canada (Green *et al.*, 2006; Arsenault *et al.*, 2012, 2013), Sweden (Nygård *et al.*, 2004), USA (Weisent *et al.*, 2011, 2012), and England and Wales (Gillespie *et al.*, 2008). With some exceptions, local studies rather deal with the mechanism of transmission to humans (Bardon *et al.*, 2009) than the spatial distribution of the infection (Zeleňáková *et al.*, 2012; Marek *et al.*, 2015a, 2015b).

The main motif of the study presented here was directly inspired by the situation mentioned above. It aims to proceed a retrospective spatio-temporal investigation with the focus on the distribution of campylobacteriosis regarding the location of fresh-milk vending machines. The study aims to discover whether any of these machines might have been the source of a local outbreak of campylobacteriosis during the period of 2008-2012, or whether most incidents could have been an expression of the disease's regular, seasonal pattern. Additional attention is drawn to the situation in České Budějovice, the regional capital of South Bohemia, that was specifically mentioned in the original announcement referred to above (Czech Republic Ministry of Health, 2010). Spatio-temporal clustering and geographic profiling were utilised to uncover the spatio-temporal patterns of campylobacteriosis.

Materials and Methods

Data

The essential dataset in the this study came from the Information System of Infectious Diseases Epidemiological Database (EPIDAT), a database covering the mandatory nationwide reporting, recording and analysis of infectious diseases in the Czech Republic (<http://www.szu.cz/publikace/data/infekce-v-cr>), managed by The National Institute of Public Health the Czech Republic (NIPH CR) and the MoH CR since 1993. EPIDAT is the basis for local, regional, national and international control of infec-

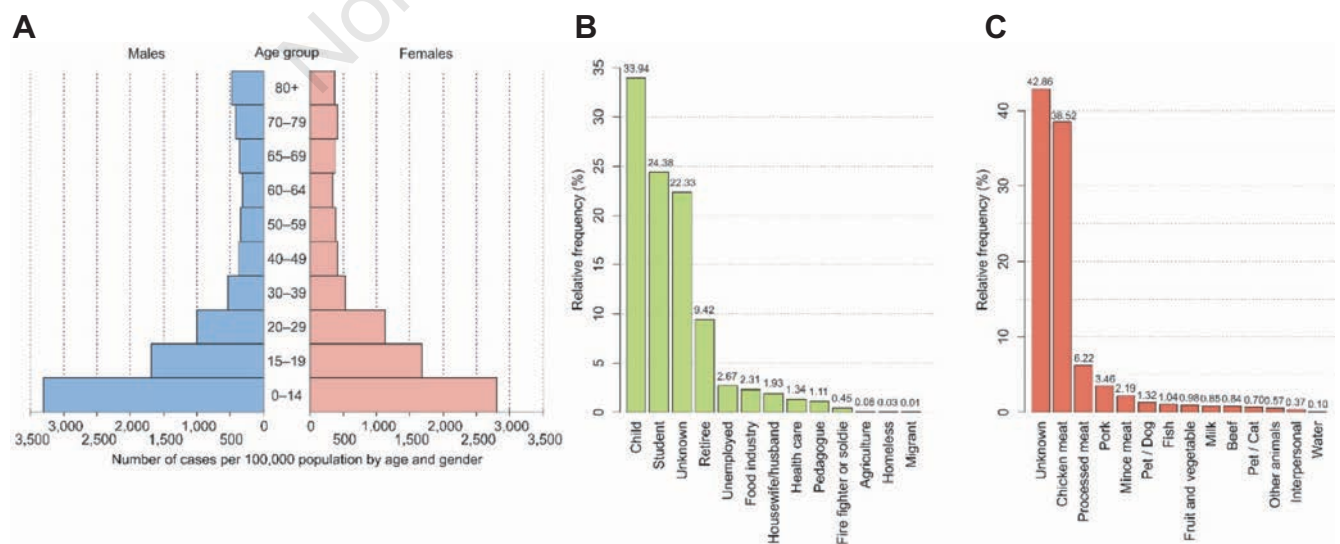


Figure 1. Characteristics of population affected by campylobacteriosis (2008-2012). Age- and gender-specific incidence (A); relative frequency of individual cases by employment (B); and relative frequency of individual cases by cause (C).



tious diseases. The data storage ensures secure exchange of actual datasets on the prevalence of infections among the departments of National Health Service (NHS), MoH CR and Czech Republic National Institute of Public Health (2016). The campylobacteriosis dataset, provided by the NIPH CR, contains facts about 98,933 recorded cases of the disease in the Czech Republic in the period 2008–2012. All records were available anonymously, so localisation of individual events could not be done better than the street level in towns and cities. The dataset also contains 78 characteristics of recorded cases such as date of record, gender, age, a potential source of infection, and occupation. However, these data vary in completeness, quality and consistency throughout the database so all records needed to be checked, cleaned, and repaired before analysis. It was possible to geocode the checked data assigning geographical coordinates to all records based on available (partial) address information. However, geocoding 98,500 records with incomplete addresses is taxing as the tools are generally limited, e.g., the Google Geocode tool API allows only 2,500 queries per day) and OpenStreetMap Nominatim is not recommended for geocoding of high volumes of data even if it can approximate addresses. The geocoding was therefore conducted using the R script querying Mapy API (<https://api.mapy.cz/>), a programme application interface from the Czech company Seznam (<https://www.seznam.cz/>).

The State Veterinary Administration of the Czech Republic (SVA CR) keeps the records about the location of all active fresh-milk vending machines. Any user can download the data from the organisation's Portal of Registered Subjects under the category of raw milk vendors (http://eagri.cz/public/app/svs_pub/subjekty/mleko.php). However, the portal stores only information on operating vending machines, so historical data cannot be obtained automatically. For that reason, we used alternative sources, such as the (now inactive) platform Countryside Forum (Venkovské Forum) and producers of vendor machines (e.g. the Toko chain), which also store these data. By the combination of these sources, a dataset consisting of 267 fresh-milk vending machines active between 2008 and 2012 was assembled.

The final dataset needed was the National Census 2011 data covering the demographic structure of administrative units, which was essential for the analyses of spatio-temporal clustering.

Space-time clusters

The time dimension is still mostly treated as one of many data characteristics during geospatial data analyses rather than emphasising its genuine inseparable part allowing the more complex overview of a situation under analysis. Main reasons for that might be limited support in common geographical information systems (GIS) software platforms coupled with higher requirements for data handling, analysis and visualisation. Due to these reasons, spatiotemporal analyses are mostly conducted in separate space and time dimensions or (better) in time slices. However, methods utilising joint spatio-temporal approach exist. One of them is spatio-temporal (permutation) scan statistics by Martin Kulldorff and implemented in SaTScan software (Kulldorff and Information Management Services Inc, 2009). The Kulldorff spatio-temporal scan statistic uses the concept of moving cylindrical windows with a circular or elliptical base where the height corresponds to time. The window changes its parameters (according to settings) while moving in space and time. This means that it permutes all possible space-time combinations theoretically creating an infinite number of overlapping cylinders, each of them representing a possible

cluster (Kulldorff *et al.*, 2005). A likelihood ratio test, which is based on the number of recorded cases and the expected number of cases in and outside the cylinder is computed for every combination, and potential clusters are identified accompanied by their significance (P value). The analysis distinguishes between high-value clusters (characterized by higher RR and more strongly affected areas) and low-values clusters (characterized by lower RR and healthier areas). Simultaneously, it also evaluates whether the statistically significant clusters represent a primary (the most likely) cluster or several secondary clusters. A benefit of spatio-temporal scan statistics, as defined by Kulldorff, is that the approach is both deterministic (identification of cluster location) and inferential (hypothesis testing and evaluation of cluster significance) (Osei, 2014). The formalised description of the process can be found in various follow-up manuscripts (Kulldorff and Nagarwalla, 1995; Kulldorff, 1997, 1999; Kulldorff *et al.*, 2005).

Spatio-temporal scan statistic using SaTScan 9.3 allows the identification of cluster patterns of administrative units, healthy or affected by campylobacteriosis and differentiates the results from random processes in the study area. The input data used consisted of gender- and age-stratified disease events aggregated spatially with regard to municipalities (city districts in the case of bigger towns and cities) at weekly intervals. The complementary data included the demographic structure of administrative units (gender and age), and their locations in the form of centroid coordinates. The spatio-temporal scanning was then based on such a stratified data and Poisson distribution of events. The analysis was set to find clusters based on a maximal size of 3% of the population in the circular window (Weisent *et al.*, 2011; Marek *et al.*, 2015b) and with a maximal temporal cluster size set at 50% (100% in case of purely spatial clusters). Nonparametric temporal trend adjustment randomization (with time stratified) was applied to ensure the comparability of rates within various periods (Kulldorff, 1999). The significance of identified clusters was assessed at P values lower than 0.05 and performed by 999 Markov chain Monte Carlo (MCMC) realisations. Then, indirectly standardised rates expressed as RR were computed for each identified geographic cluster according to Green *et al.* (2006) so that only significant clusters remained in the output files. Final visualisation was created using ArcGIS 10.3 (ESRI, Redlands, CA, USA).

Geographic profiling

Geoprofiling is a method originally developed for the statistical evaluation in criminology, where it has been successfully utilised for identification of the most likely home addresses of crime offenders based on serial delinquencies (Stevenson *et al.*, 2012). Recently, the method has also been applied to tasks disparate from its original purpose, such as in field of species modelling (Martin *et al.*, 2009; Raine *et al.*, 2009) or spatial epidemiology (Buscema *et al.*, 2009; Le Comber *et al.*, 2011), where single or multiple sources (nests, feeding places, outbreaks, *etc.*) may be identified. The technique uncovers common hidden geographical patterns of events and/or similar behaviour. These can be regular movements in a familiar area, similar feeding patterns or a common source in disease outbreaks. In addition, a sufficient amount of events allows the construction of the most likely estimate (Snook *et al.*, 2005). Geoprofiling seeks original locations of events overcoming the methodological limitation of spatially aggregated data by revealing continuous patterns and processes that might be masked, e.g. when using administrative units as map inputs (Le Comber *et al.*, 2011). The performance of a geoprofile

is determined by a measure called the hit score percentage (HS%). This is defined as the ratio of the size of an area to be searched following geoprofile prioritisation compared to the total area in question before a specific address is found (Rossmo, 2014). A HS% of 50% (0.5) represents either random or regular search patterns, while a HS% of 2% (0.02) is used in most of studies to distinguish the most likely geoprofile (Rossmo, 1999).

The research presented here utilised a modification of geoprofiling based on the usage of the Dirichlet process mixture (DPM) model, which benefits from Bayesian modelling of clusters using the MCMC application (Verity *et al.*, 2014). Since DPM does not require an estimation of a number of clusters beforehand (*i.e.* the number of sources), it is suitable for applications where the number of clusters is unknown or impossible to assume. The mathematical apparatus and analytical solution of DPM have been described by Verity *et al.* (2014).

Geoprofiling was applied to the geocoded records of campylobacteriosis cases together with the locations of fresh-milk vending machines in the Czech Republic in the 2008–2012 period. The number of disease events was reduced to cases that appeared in the area up to 15 km to the nearest vending machine, which not only considers local people but also those commuting to work or to school (Horák and Ivan, 2010). By setting this distance threshold, the data were reduced to 50,000 records subdivided into 18 regions. This approach reduced the computationally demanding MCMC implementation of geoprofiling and facilitated catching the local character of analysis and also

The geoprofile was created using Rgeoprofile (Stevenson and Verity, 2014; R Core Team, 2016). The generation of geoprofiles using Rgeoprofile is straightforward; firstly data are loaded, then parameters for graphics, model and simulation are set, and then simulation commences. The most important parameter is the

Sigma defining the distance-decay function, and that was set to 0.1° (approximately 1 km in the study area). Thinning is of importance in MCMC sampling parameters setup as it defines the reduction of the data. Its level is estimated through the autocorrelation plot as depicted in Figure 2 for South Bohemia. Then the geoprofile with HS%*s* was constructed and visualised. The number of disease events in study area subdivisions, the number of vending machines in the subdivision, the time of analysis, and the level of thinning together with a number of potential outbreak sources are mentioned in Table 1. The visualisation of the final geoprofile was created using ArcGIS 10.3.

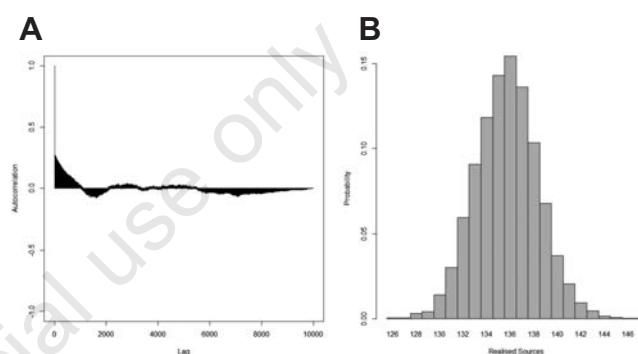


Figure 2. Autocorrelation plot of simulations (A) and number of realised sources (B) in South Bohemia.

Table 1. Study area subdivisions, geoprofile settings and potential sources identified.

Area	Cases of campylobacteriosis (n)	Thinning	Vending machines (n)	Potential sources (n)	Time for analysis (h)
Brno	6,116	100	23	4	36.00
Hana (Hanakia)	3,180	100	13	3	11.13
South Bohemia	3,046	1,000	26	6	9.63
South Moravia	1,494	100	11	5	2.85
Opava	1,599	100	6	2	3.03
Ostrava	4,657	1,500	8	0	13.30
Plze (Pilsen)	3,082	500	24	8	7.90
Prague	4,828	600	13	1	26.94
North Bohemia	694	100	9	3	0.44
North Moravia	1,214	100	14	5	1.36
East Bohemia	3,985	100	27	1	19.15
Silesia	3,052	1,200	16	1	6.07
Slovácko*	3,472	100	28	4	13.12
Central Bohemia-east	1,276	100	3	0	1.56
Central Bohemia-west	2,010	2,000	7	0	3.53
Valašsko (Wallachia)	3,528	100	14	1	13.57
Vysočina Region	711	100	6	0	0.52
West Moravia	2,088	50	19	7	5.10
Sum	50,032		267	51	175.19

*Moravian Slovakia.

Results

Spatio-temporal clustering

The scanning performed allowed the identification of 30 statistically significant clusters ($P < 0.001$) during the study period with the locations depicted in Figure 3. Fourteen clusters were high-value clusters consisting of areas more prone to occurrence of campylobacteriosis indicated by a strongly increased RR (> 1.50). Details are given in Table 2. The primary (and most likely) cluster ($RR = 2.16$) was located in the north-east of the Czech Republic in the city of Ostrava and surroundings and consisted of 31 municipalities and city districts with a population of almost 293,000 (indicated as 1 in Figure 3 and Table 2). Although the other clusters were denoted as secondary, they were still considered statistically significant as well as locally important. High-rate clusters no. 2 and 3 were closely connected to the primary cluster and together with cluster no. 4, they were also the most homogenous. On the contrary, the most heterogeneous ones were mainly the large-scale clusters no. 5, 9, 10 and 13. Only six clusters had average RRs equal to or less than 2.00.

Most of the high-rate clusters (9 out of 14) were located in Moravia and Silesia (in the eastern part of the Czech Republic),

while five of them were found in Bohemia (in the western and central parts). Most of these clusters were also present throughout the study period, only five of them (no. 5, 10, 11, 13, 14) appeared only at specific times. Clusters no. 5, 13 and 14 occurred in the warm part of the year, while cluster no. 10 (in south Wallachia) covered more than 1.5-years. The cluster with the overall highest RR (5.41, no. 11) appeared in the city of České Budějovice and its neighbourhood. It consisted of up to 60 municipalities affecting almost 158,000 people but it was also the high-rate cluster of the shortest duration, *i.e.* six weeks in the winter of 2010. This result was noted as an exceptional situation and later commented on by the relevant authorities. Up to 25% (2.6 million) of the Czech population lived in areas classified as high-rates clusters during the 2008-2012 period.

Sixteen low-rate clusters ($RR \leq 0.80$) were identified during the study period, all of them in Bohemia. The first group appeared in mountainous/highland municipalities with a relatively low population densities (no. 15, 17, 20, 21, 22, 25, and 26). The second group appeared in rather densely populated areas on the outskirts of mid-sized municipalities. Only two of the low-rate clusters (no. 26 and 29) were temporally specific, lasting for five and 19 months, respectively. Up to 37% (3.9 million) of the Czech population lived in areas classified as low rates clusters during the study period.

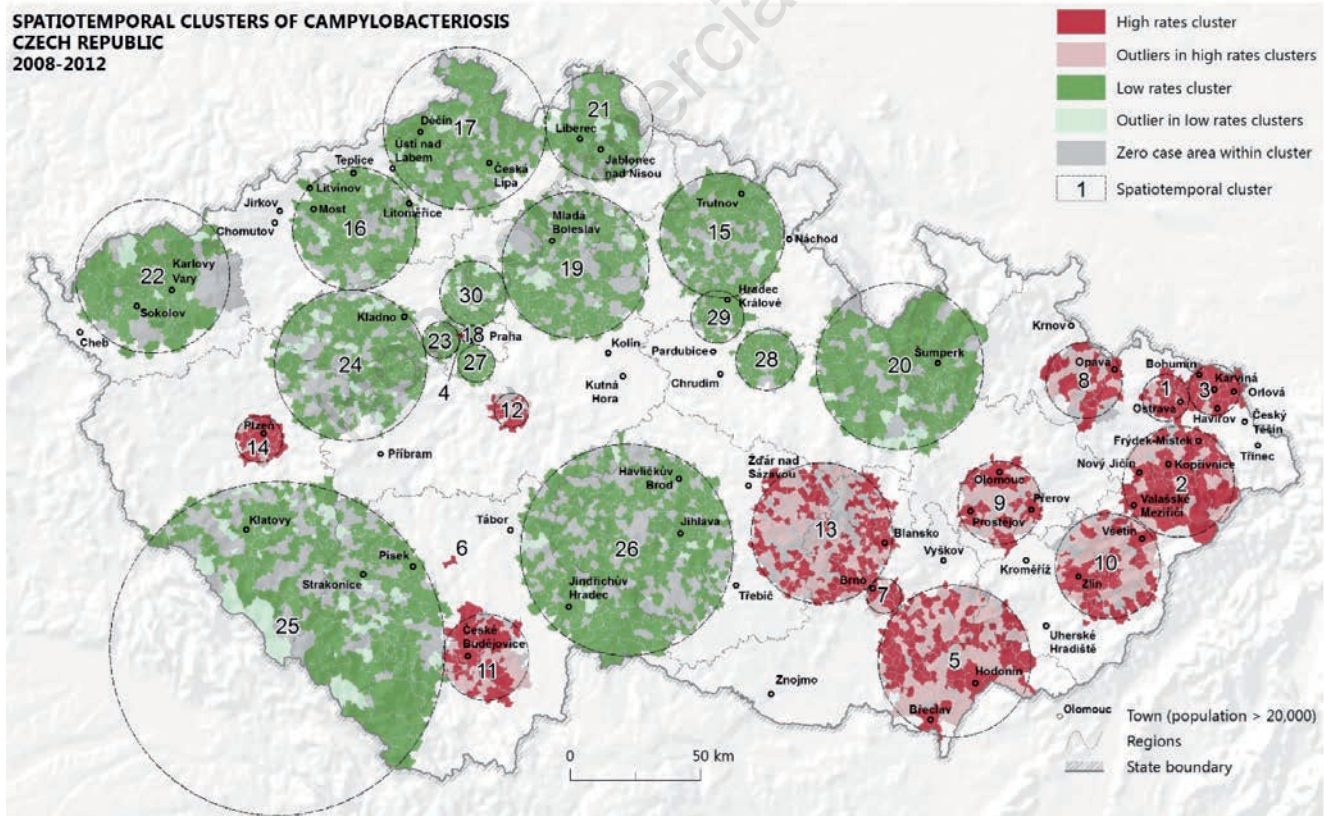


Figure 3. Location of spatio-temporal clusters identified in the Czech Republic in the period 2008-2012 using SaTScan. Red=affected areas represented by high-value spatiotemporal clusters ($RR > 1.50$); green=healthier areas represented by low-value spatiotemporal clusters ($RR \leq 0.80$). The clusters were heterogeneous with lighter shades indicating less strongly affected areas and grey municipalities without disease occurrence.

Table 2. Spatiotemporal clusters of high values identified in the Czech Republic in the period 2008-2012.

Cluster	Time (from-to)	Region in the Czech Republik	Municipalities in cluster (n)	Observed cases in the cluster (n)	Expected cases in the cluster (n)	RR	Population
1	2008/01/01-2012/12/31	Ostrava	31	5975	2861	2.16	292,978
2	2008/01/01-2012/12/31	North Wallachia - Lachia	70	5414	2788	2.00	277,236
3	2008/01/01-2012/12/31	Silesia	16	4773	2534	1.93	256,657
4	2008/01/01-2012/12/31	Prague - centre	1	1006	245	4.13	29,948
5	2008/05/13-2010/11/01	South Moravia	167	2274	1432	1.60	292,885
6	2008/01/01-2012/12/31	Dražič	1	7	2	4.19	214
7	2008/01/01-2012/12/31	Brno - centre	19	3951	2590	1.55	271,742
8	2008/01/01-2012/12/31	Opava	37	1714	877	1.97	87,203
9	2008/01/01-2012/12/31	Haná	66	3828	2526	1.54	256,721
10	2009/04/14-2011/09/05	South Wallachia	90	1596	932	1.72	196,522
11	2010/01/12-2010/02/22	České Budějovice	60	194	36	5.41	157,425
12	2008/01/01-2012/12/31	Benešov	15	640	313	2.05	31,115
13	2010/04/06-2010/10/04	Brno - surroundings	224	568	286	1.99	284,346
14	2011/05/03-2011/11/14	Pilsen	22	394	201	1.96	197,263

RR, relative risk. P value of all clusters is <0.001.

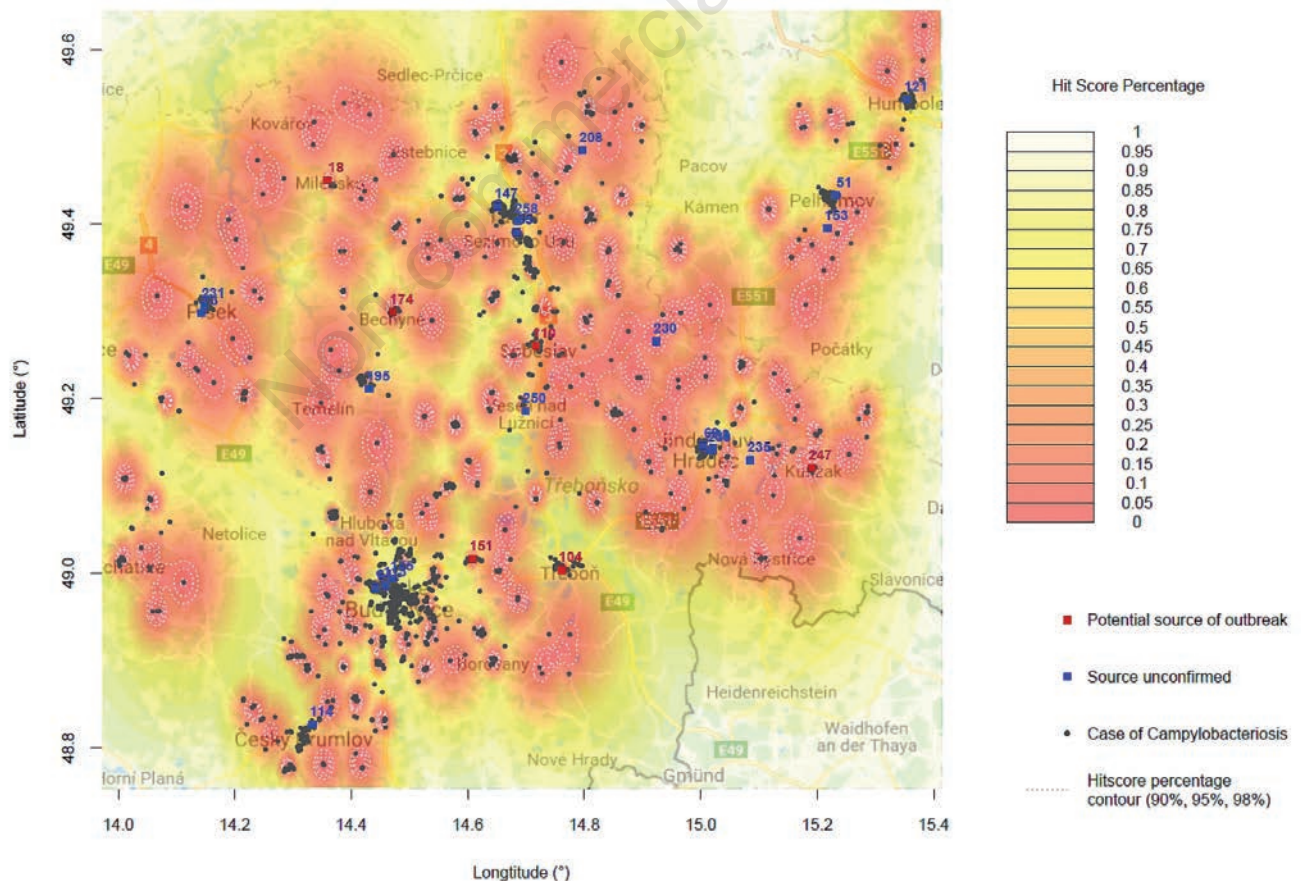


Figure 4. Geographical profile constructed for South Bohemia (created by the adapted functionality of the Rgeoprofile package).

Geographical profiling

The geoprofiling approach was utilised to find out which fresh-milk vending machines could have been sources of campylobacteriosis outbreaks. All subdivisions of the study area were evaluated by individual geoprofiles presenting HS% surfaces that were combined with the locations of vending machines in the sub-area. Initially, the analyses identified 51 machines (19.1%) as potential sources of campylobacteriosis in the 2008-2012 period. The geoprofile constructed for South Bohemia (Figure 4) identified six fresh-milk vending machines that could have caused some of the local outbreaks. However, none of them was located in the city of České Budějovice, the place where the risk argumentation had begun. Although two potential sources were found in the two nearby towns Lišov (12 km/15 min) and Třeboň (26 km/30 min), the vending machines there were installed at later dates: beginning of February 2010 in Třeboň and in August 2010 in Lišov. However, after additional control, it was found that the DPM had overrated the HS% with regard to those sources having only small numbers of cases in their vicinity. Based on this finding, the potential of individual vending machines to be outbreak sources was categorised as (1) an unconfirmed source; (2) a potential source with less than 30 cases within 2 km radius; and (3) a potential source of 30 cases or more within 2 km radius. By this procedure, only 24 fresh milk vending machines (9%) were considered to be the more likely potential source of an outbreak in the area.

All geoprofiles were combined into a surface covering the whole of the Czech Republic, visualizing the combined HS% surfaces and location of the vending machines (Figure 5). The map also differentiates between areas within 15 km distance from a vending machine (depicted in saturated colours) with further areas masked. Since the final surface was created from combined geoprofiles, one can notice the different size of thinning set up recognisable by the size of the lowest HS% hotspots.

Based on the geoprofiling analysis, the Czech Republic can be split into four zones concerning the outcome of the evaluation of the individual vending machines (for the location of zones see the small map at top left corner of Figure 5). The first zone covers the southern part of Bohemia where the highest ratio of machines as a potential source of a local outbreak with less than 30 cases were identified. It also had the lowest density of vending machines (2.2 per 1,000 km²). Northern Bohemia (Zone 2) contained the least percentage (and also number) of potential vending machine sources of the local outbreak. Zone 3 covers the western part of Moravia and Silesia and had the highest percentage of potential vending machine sources (20.6%) of local outbreaks with more than 30 cases in their proximity. Zone 4 is located in eastern Moravia and Silesia and was conspicuous with the highest number of unconfirmed sources, the highest density of vending machines (7.1 per 1,000 km²) and still the second lowest number of identified sources with more than 30 cases.

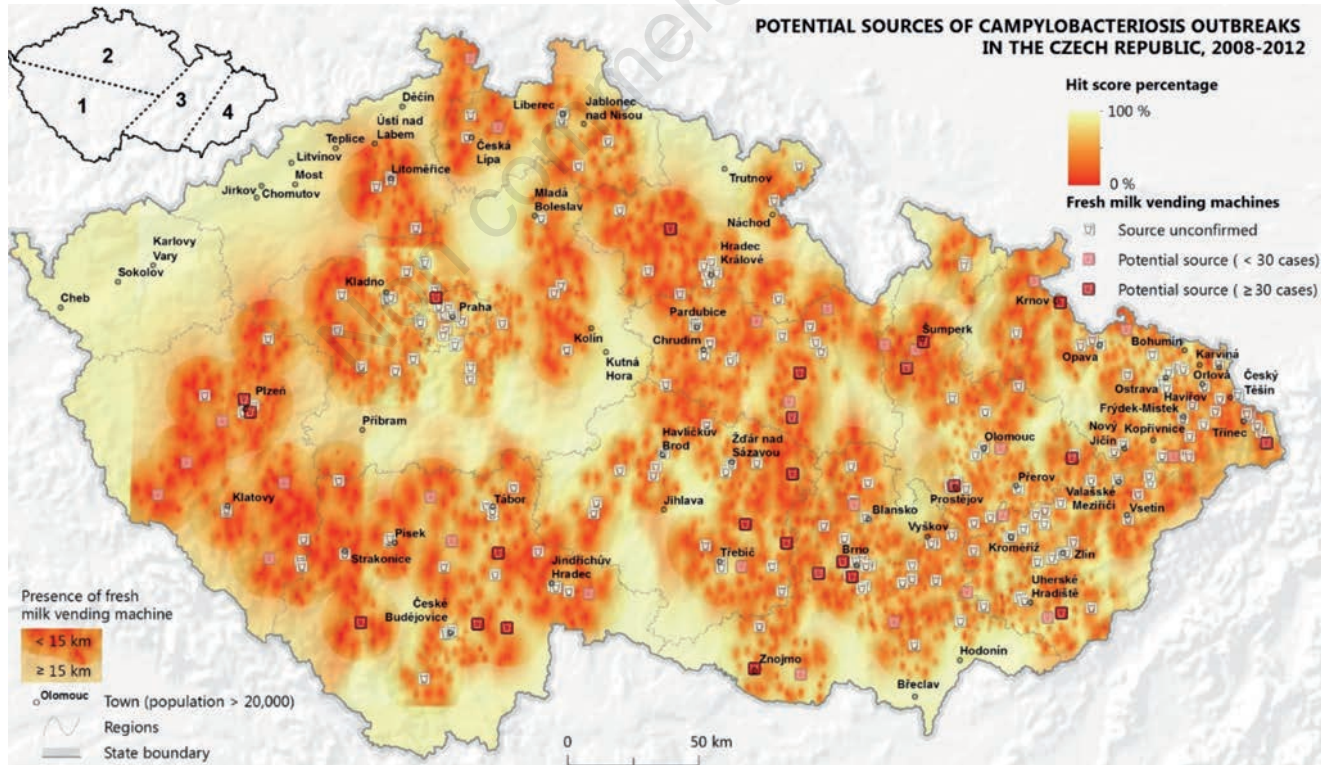


Figure 5. Combined geoprofile for the evaluation of fresh-milk vending machines as potential sources of campylobacteriosis outbreaks in the Czech Republic, 2008-2012. High likelihood of finding an outbreak source are shown in strong colour; light yellow and orange shades indicate unlikely areas; vending machines unconfirmed as an outbreak source are shown as white squares; possible sources with less than 30 cases in the vicinity are shown in a light shades of red, while the most likely sources are shown in red.

Discussion

Statements about the lack of relevant health datasets commonly leave out the whole truth. Rather than exclusively focusing on whether or not they exist, the question should also include the availability of the data and the conditions that rule access to them. Even though the geocoding process resulted in satisfactory precise locations for most cases of campylobacteriosis in our case, one must be aware of positioning uncertainty and inaccuracies that are primarily caused by the anonymity of the original data. Furthermore, the location of each case is mostly given by the (anonymized) residential address of the patient rather than the place of infection. However, localised data could be utilised in spatial, temporal or spatio-temporal statistical analyses with a suitable accuracy, either as spatial points or in the form of aggregated counts within administrative units. The selection of suitable units for aggregation is crucial due to the ecological fallacy (Jeffery *et al.*, 2014) or modifiable areal unit problem (MAUP) (Openshaw, 1984). Results originating from a certain level of spatial aggregation should not be generalised at different scales, or even at the individual level and the issue is then to find a reasonable compromise among the maximal benefit of local analysis and data confidentiality.

The successful analysis is dependent on the inclusion of suitable parameter settings in the spatio-temporal scan statistic. Apart from the results presented above, combinations of other settings, including spatial, temporal, and analytical parameters, were tested, *e.g.*, the ratio of the affected population (3%, 5%, 10% and 50%), the maximal period of clusters existence (30 or 105 days or 50% of the study period) and temporal trend analysis. However, the results did not change considerably under any such parameter changes, except that less larger and more populated clusters were identified as the result of aggregation, meaning the same population in the risk nonetheless hiding the local variations. The primary cluster was always found in the same location and with similar characteristics. The final settings of parameters were selected experimentally with regard to findings of previous, related studies (McLaughlin and Boscoe, 2007; Chen *et al.*, 2008; Yiannakoulias, 2009; Weisent *et al.*, 2011).

Although geoprofiling is not a genuine method for the study of spatial clustering, it can be used to examine the spatial pattern of events. Originally used in criminology, it has also found use in spatial epidemiology, ecology and biology. Geoprofiling assumes that the dynamics of analysed phenomena show similar patterns of spatial behaviour that can therefore easily be mapped. Based on the geoprofile, it is possible to estimate the source of the outbreak from a set of candidate places that can be evaluated by the likelihood surface created by the DPM model from event points entered. Hence, the method takes into account only the location of individual events and no other attributes or even time. Possible variances

in estimates noted could have been caused by changes in the settings in the R simulation or by the definition of distances and settings of thinning, mostly based on subjective evaluation of autocorrelation plot coming from MCMC simulations. Description of the spatial pattern of populations, events or behaviour could also be inadequate. In this connection, it is sufficient to remember that the spatial accuracy of geodata created by geocoding from data do not reach further down than the street level. Another limitation of geoprofiling as used in the analysis presented here is that the method is exclusively spatial, which produces results for the entire study period rather than any specific time. One way to prevent this would be to utilise time slices of geodata.

Geoprofiling initially identified 19% of vending machines as potential sources of local campylobacteriosis outbreaks. However, after further exploration, it was noted that the method favoured and overrated the situation concerning some of the vending machines, *i.e.* those with very low numbers of cases in their vicinity. Further adjustment by the number of events in the nearest neighbourhood of identified sources allowed correcting this number to 9%. It should also be mentioned that geoprofiling does not serve to confirm outbreak sources but tries to select them based on spatial patterns similar to the phenomenon explored. The fact that certain vending machines were selected as potential sources of outbreaks does not necessarily mean that this indeed was the real source since the origin was unknown in almost 43% of recorded cases in the Czech Republic during the study period. Therefore, the presented geoprofile for the area of the Czech Republic describes a probable rather than exact scenario of the situation in 2008-2012.

Previous studies (Marek *et al.*, 2015a, 2015b) show increased prevalence rates of campylobacteriosis in the eastern part of Moravia, the same area that also has the highest number (and density) of vending machines. However, there was also a second lowest percentage area, and a number of vending machines detected as a potential source (Table 3). Thus, the direct connection between local outbreaks and fresh-milk vending machines seems rather weak. The machines are also a subject of strict hygienic standards and regular testing of the milk. Although the results of tests would help to verify the results of our analyses, they are not publicly available.

The limitation of this study is the focus on spatial and spatio-temporal patterns of *campylobacter* human infections in the Czech Republic, which overshadows other possible geographical or socio-economic determinants. Examples include ruminant and poultry density (Bessell *et al.*, 2010), demographic structure (Ethelberg *et al.*, 2005; Gabriel *et al.*, 2010), food outlets location and food consumption characteristics (Rind and Pearce, 2010) and the climate (Kovats *et al.*, 2004). A direct comparison with similar studies in the Czech Republic is also omitted since these were mostly focused on the aetiology of *Campylobacter spp.* infections in the context of biomedical, and veterinary science (Videnska *et*

Table 3. Distribution of fresh-milk vending machines in the Czech Republic along with the estimated risk of infection source.

Zone	Vending machines (n)	Unconfirmed sources (n)	Potential sources with <30 cases (n)	Potential sources with ≥30 cases (n)
1=Southern Bohemia	49	35	8	6
2=Northern Bohemia	71	64	5	2
3=Western Moravia and Silesia	63	42	8	13
4=Eastern Moravia and Silesia	84	75	6	3



al., 2014) or microbiology (Steinhauserova *et al.*, 2002). However, Marek (2015) analysed associations among the RR of the disease and a number of possible determinants, from which population density, temperature and the impact of agricultural/meat-processing businesses in individual municipalities were identified as the most influential ones, so their inclusion could potentially have improved the presented analyses. The uniqueness of the study resides in the combination of geoprofiling, a fine-scale (street) level of approach and the spatial extent covering the area of the Czech Republic as a whole.

Conclusions

The complex aetiology of campylobacteriosis is still not fully discovered, while it has been recorded that even some dairy products, although testing negatively for *Campylobacter spp.*, can be a source of infection by the disease (Center for Disease Control and Prevention, 2009).

Spatial and spatio-temporal techniques have recently become common and have found usage in the health-related research. Spatio-temporal scanning is one of these widely used tools that allow us to explore spatio-temporal patterns of epidemiological events and phenomena. The combination of these approaches with geoprofiling has made it possible to progress further in search for unknown sources of infection.

Despite fresh-milk vending machines being subject to strict hygiene standards and regular monitoring, a potential link and the spatial distribution of campylobacteriosis has been detected in the Czech Republic. This study was inspired by this situation in the South Bohemia region around the city of České Budějovice. No vending machines were detected by geoprofiling as potential outbreak source in the city itself, although a significant spatio-temporal cluster of high rates was found affecting a large number of people at the beginning of 2010. Two potential sources were identified in some nearby municipalities but vending machines there were installed after the start of the outbreak.

References

- Agrarian Chamber of the Czech Republic, 2010. Bogeyman called fresh milk vending machine (in Czech). Available from: <http://www.apic-ak.cz/strasak-zvany-automat-na-cerstve-mleko.php>
- Ambrožová H, 2011. Summer diarrhoeas (in Czech). *Medicína Pro Praxi* 8:7-9.
- Andrlová L, 2011. The perspectives of selling (organic) milk in milk vending machines (in Czech). University of South Bohemia, Ceske Budejovice, Czech Republic.
- Arsenault J, Berke O, Michel P, Ravel A, Gosselin P, 2012. Environmental and demographic risk factors for campylobacteriosis: do various geographical scales tell the same story? *BMC Infect Dis* 12:1-12.
- Arsenault J, Michel P, Berke O, Ravel A, Gosselin P, 2013. How to choose geographical units in ecological studies: Proposal and application to campylobacteriosis. *Spat Spatiotemporal Epidemiol* 7:11-24.
- Bardon J, Kolar M, Cekanova L, Hejnar P, Koukalova D, 2009. Prevalence of *Campylobacter jejuni* and its resistance to antibiotics in poultry in the Czech Republic. *Zoonoses Public Health* 56:111-6.
- Bessell PR, Matthews L, Smith-Palmer A, Rotariu O, Strachan NJC, Forbes KJ, Cowden JM, Reid SWJ, Innocent GT, 2010. Geographic determinants of reported human *Campylobacter* infections in Scotland. *BMC Public Health* 10:1-8.
- Buscema M, Grossi E, Breda M, Jefferson T, 2009. Outbreaks source: A new mathematical approach to identify their possible location. *Phys A Stat Mech* 388:4736-62.
- Center for Food Security & Public Health, 2013. Campylobacteriosis. Iowa State University, Ames, IA, USA.
- Center for Disease Control and Prevention, 2009. *Campylobacter jejuni* infection associated with unpasteurized milk and cheese - Kansas, 2007. Available from: <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5751a2.htm>
- Chen J, Roth RE, Naito AT, Lengerich EJ, Maceachren AM, 2008. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *Int J Health Geogr* 7:57.
- Czech Republic Ministry of Health, 2010. Ministry of Health alerts about possible health risks caused by the consumption of raw milk. Ministry of Health of the Czech Republic, Prague, Czech Republic.
- Czech Republic National Institute of Public Health, 2016. Infections in the Czech Republic -EPIDAT (in Czech). Available from: <http://www.szu.cz/publikace/data/infekce-v-cr>
- Domašová I, 2014. Campylobacteriosis: Report of Regional Hygienic Station (in Czech), Regional Hygiene Station of Central Bohemian Region, Prague, Czech Republic.
- Ekdahl K, Normann B, Andersson Y, 2005. Could flies explain the elusive epidemiology of campylobacteriosis? *BMC Infect Dis* 5:1-4.
- Ethelberg S, Simonsen J, Gerner-Smidt P, Olsen KEP, Mølbak K, 2005. Spatial distribution and registry-based case-control analysis of *Campylobacter* infections in Denmark, 1991-2001. *Am J Epidemiol* 162:1008-15.
- Gabriel E, Wilson D, Leatherbarrow A, Cheesbrough J, Gee S, Bolton E, Fox A, Fearnhead P, Hart C, Diggle P, 2010. Spatio-temporal epidemiology of *Campylobacter jejuni* enteritis, in an area of Northwest England, 2000-2002. *Epidemiol Infect* 138:1384-90.
- Gillespie I, O'Brien SJ, Penman C, Tompkins D, Cowden J, Humphrey TJ, 2008. Demographic determinants for *Campylobacter* infection in England and Wales: implications for future epidemiological studies. *Epidemiol Infect* 136:1717-25.
- Green CG, Krause DO, Wylie JL, 2006. Spatial analysis of campylobacter infection in the Canadian province of Manitoba. *Int J Health Geogr* 5:1-14.
- Havelaar A, Ivarsson S, Löfdahl M, Nauta M, 2013. Estimating the true incidence of campylobacteriosis and salmonellosis in the European Union, 2009. *Epidemiol Infect* 141:293-302.
- Horák J, Ivan I, 2010. Transport networks and their influence to potential commuting in the Czech Republic. *Životné Prostr* 44:159-67.
- Hygiena, 2010. The safety to health of milk distributed in vending machines (in Czech). *Hygiena* 55:34.
- Institute of Health Information and Statistics of the Czech Republic, 2013. Infectious Diseases in 2013 (in Czech), Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic.
- Jagai JS, Castronovo DA, Naumova EN, 2007. The use of Köppen

- climate classification system for public health research. *Epidemiology* 18:7-212.
- Jeffery C, Ozonoff A, Pagano M, 2014. The effect of spatial aggregation on performance when mapping a risk of disease. *Int J Health Geogr* 13:1-9.
- Kovats RS, Edwards SJ, Hajat S, Armstrong BG, Ebi KL, Menne B, 2004. The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. *Epidemiol Infect* 132:443-53.
- Kulldorff M, 1997. A spatial scan statistic. *Commun Stat Theory Method* 26:1481-96.
- Kulldorff M, 1999. Spatial scan statistics: models, calculations, and applications. In: Glaz J, Balakrishnan N (eds.) *Scan statistics and applications*. Birkhäuser, Boston, MA, USA, pp. 303-22.
- Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F, 2005. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* 2:e59.
- Kulldorff M, Information Management Services Inc, 2009. SaTScan v9.3: Software for the spatial and space-time scan statistics. StatScan, Boston, MA, USA.
- Kulldorff M, Nagarwalla N, 1995. Spatial disease clusters: detection and inference. *Stat Med* 14:799-810.
- Le Comber SC, Rossmo DK, Hassan AN, Fuller DO, Beier JC, 2011. Geographic profiling as a novel spatial tool for targeting infectious disease control. *Int J Health Geogr* 10:1-8.
- LeJeune JT, Rajala-Schultz PJ, 2009. Unpasteurized milk: a continued public health threat. *Clin Infect Dis* 48:93-100.
- Lexová P, Beneš Č, Částková J, Kynčl J, Králová R, 2013. Foodborne and waterborne infections in the Czech Republic in 2012 and incidents trends (in Czech). *Zprávy CEM* 22:233-39.
- Marek L, 2015. Spatial and multivariate statistical analyses of epidemiological data (in Czech). Palacky University, Olomouc, Czech Republic.
- Marek L, Pászto V, Tuček P, 2015a. Bayesian mapping of medical data. In: Brus J, Vondrakova A, Vozenilek V (eds.) *Modern trends in cartography*. Springer International Publishing, Berlin, Germany, pp. 489-505.
- Marek L, Tuček P, Pászto V, 2015b. Using geovisual analytics in Google Earth to understand disease distribution: a case study of campylobacteriosis in the Czech Republic. *Int J Health Geogr* 14:1-13.
- Martin R, Rossmo DK, Hammerschlag N, 2009. Hunting patterns and geographic profiling of white shark predation. *J Zool* 279:111-8.
- McLaughlin CC, Boscoe FP, 2007. Effects of randomization methods on statistical inference in disease cluster detection. *Health Place* 13:152-63.
- Mullner P, Shadbolt T, Collins-Emerson JM, Midwinter C, Spencer SEF, Marshall J, Carter PE, Campbell DM, Wilson DJ, Hathaway S, Pirie R, French NP, 2010. Molecular and spatial epidemiology of human campylobacteriosis: source association and genotype-related risk factors. *Epidemiol Infect* 138:1372-83.
- Nygård K, Andersson Y, Röttingen JA, Svensson A, Lindbäck J, Kistemann T, Giesecke J, 2004. Association between environmental risk factors and campylobacter infections in Sweden. *Epidemiol Infect* 132:317-25.
- Nylen G, Dunstan F, Palmer SR, Andersson Y, Bager F, Cowden J, Feierl G, Galloway Y, Kapperud G, Megraud F, Molbak K, Petersen LR, Ruutu P, 2002. The seasonal distribution of campylobacter infection in nine European countries and New Zealand. *Epidemiol Infect* 128:383-90.
- Openshaw S, 1984. *The modifiable areal unit problem, concepts*. GeoBooks, Norwich, UK.
- Osei FB, 2014. Current statistical methods for spatial epidemiology: a review. *Austin Biometr Biostat* 1:1-7.
- R Core Team, 2016. R: A language and environment for statistical computing. Available from: <https://www.r-project.org/>
- Raine NE, Rossmo DK, Le Comber SC, 2009. Geographic profiling applied to testing models of bumble-bee foraging. *J R Soc Interface* 6:307-19.
- Rind E, Pearce J, 2010. The spatial distribution of campylobacteriosis in New Zealand, 1997-2005. *Epidemiol Infect* 138:1359-71.
- Rossmo DK, 1999. *Geographic profiling*. CRC Press, Boca Raton, FL, USA.
- Rossmo DK, 2014. Geographic profiling. In: *Encyclopedia of criminology and criminal justice*. Springer, New York, NY, pp. 1934-42.
- Sari Kovats R, Edwards SJ, Charron D, Cowden J, D'Souza RM, Ebi KL, Gauci C, Gerner-Smidt P, Hajat S, Hales S, Hernández Pezzi G, Kriz B, Kutsar K, McKeown P, Mellou K, Menne B, O'Brien S, Van Pelt W, Schmid H, 2005. Climate variability and campylobacter infection: An international study. *Int J Biometeorol* 49:207-14.
- Snook B, Taylor PJ, Bennell C, 2005. Commentary: shortcuts to geographic profiling success: A reply to Rossmo (2005). *Appl Cogn Psychol* 19:655-61.
- Steinhausova I, Ceskova J, Nebola M, 2002. PCR/restriction fragment length polymorphism (RFLP) typing of human and poultry *Campylobacter jejuni* strains. *Lett Appl Microbiol* 34:354-8.
- Stevenson MD, Rossmo DK, Knell RJ, Le Comber SC, 2012. Geographic profiling as a novel spatial tool for targeting the control of invasive species. *Ecography (Cop)* 35:704-15.
- Verity R, Stevenson MD, Rossmo DK, Nichols R, Le Comber SC, 2014. Spatial targeting of infectious disease control: identifying multiple, unknown sources. *Methods Ecol Evol* 5:647-55.
- Videnska P, Rahman MM, Faldynova M, Babak V, Matulova ME, Prukner-Radovic E, Krizek I, Smole-Mozina S, Kovac J, Szmolka A, Nagy B, Sedlar K, Cejkova D, Rychlik I, 2014. Characterization of egg laying hen and broiler fecal microbiota in poultry farms in Croatia, Czech Republic, Hungary and Slovenia. *PLoS One* 9:e110076.
- Weisent J, Rohrbach B, Dunn JR, Odoi A, 2011. Detection of high risk campylobacteriosis clusters at three geographic levels. *Geospat Health* 6:65-76.
- Weisent J, Rohrbach B, Dunn JR, Odoi A, 2012. Socioeconomic determinants of geographic disparities in campylobacteriosis risk: a comparison of global and local modeling approaches. *Int J Health Geogr* 11:1-16.
- Yiannakoulis N, 2009. Using population attributable risk to understand geographic disease clusters. *Health Place* 15:1142-8.
- Zelenáková L, Žiarovská J, Kráčmar S, Mura L, Kozelová D, Lopašovský Ľ, Kunová S, Tináková K, 2012. Application of epidemiological information system (EPIS) in the Slovak republic within the surveillance of salmonellosis and campylobacteriosis outbreaks in the European Union (2001-2010). *Acta Univ Agr Silvic Mendel Brun* 40:189-200.