Detection of high risk campylobacteriosis clusters at three geographic levels

Jennifer Weisent[1], Barton Rohrbach[1], John R. Dunn[2], Agricola Odoi[1]

[1]*Department of Comparative and Experimental Medicine, College of Veterinary Medicine, The University of Tennessee, Knoxville TN 37996, USA; [2]Tennessee Department of Health, Communicable and Environmental Disease Service, Nashville, TN 37243, USA*

Abstract. Campylobacteriosis is a leading cause of bacterial gastroenteritis in the United States and many other developed countries. Understanding the spatial distribution of this disease and identifying high-risk areas is vital to focus resources for prevention and control measures. In addition, determining the appropriate scale for geographical analysis of surveillance data is an area of concern to epidemiologists and public health officials. The purpose of this study was to (i) compare standardized risk estimates for campylobacteriosis in Tennessee over three distinct geographical scales (census tract, zip code and county subdivision), and (ii) identify and investigate high-risk spatial clustering of campylobacteriosis at the three geographical scales to determine if clustering is scale dependent. Significant high risk clusters (P <0.05) were detected at all three spatial scales. There were overlaps in regions of high-risk and clusters at all three geographic levels. At the census tract level, spatial analysis identified smaller clusters of finer resolution and detected more clusters than the other two levels. However, data aggregation at zip code or county subdivision yielded similar findings. The importance of this line of research is to create a framework whereby economically efficient disease control strategies become more attainable through improved geographical precision and risk detection. Accurate identification of disease clusters for campylobacteriosis can enable public health personnel to focus scarce resources towards prevention and control programmes on the most at-risk populations. Consistent results at multiple spatial levels highlight the robustness of the geospatial techniques utilized in this study. Furthermore, analyses at the zip code and county subdivision levels can be useful when address level information (finer resolution data) are not available. These procedures may also be used to help identify regionally specific risk factors for campylobacteriosis.

Keywords: cluster detection, spatial smoothing, *Campylobacter*, USA.

Introduction

Campylobacter are motile, spiral shaped, gram negative bacteria found ubiquitously in the environment (Snelling et al., 2005; Humphrey et al., 2007). The organisms have been identified as a leading cause of human gastroenteritis in developed nations (Allos and Taylor, 1998; Altekruse and Swerdlow, 2002). *Campylobacter* is estimated to cause illness in over 1.3 million people in the United States annually, resulting in approximately 13,000 hospitalizations (Scallan et al., 2011). *Campylobacter* species are found in a wide variety of sources and can be isolated from foods, sewage and water. As a result, causative associations between sporadic human infections and environmental sources are difficult to ascertain (Frost, 2001). The

Corresponding author:
Jennifer Weisent
Department of Comparative and Experimental Medicine
College of Veterinary Medicine
The University of Tennessee
205A River Drive, Knoxville TN 37996, USA
Tel. +1 865 406 1510; Fax +1 865 974 5640
Email: jweisent@utk.edu

risk of disease has been linked to a complex set of determinants, including geographical region, climate patterns, human behaviour, and recreational water and land use (Hearnden et al., 2003; Kovats et al., 2005; Jepsen et al., 2009). While the incidence of campylobacteriosis varies seasonally and geographically, the reasons behind this variation are not clearly understood (Nylen et al., 2002; Weisent et al., 2010). In the United States, Foodborne Diseases Active Surveillance Network (FoodNet) identifies marked geographical variations in the incidence of *Campylobacter* infection. Incidence also differs considerably between countries based on health care seeking behaviour and laboratory culturing practices (Vally et al., 2009). In 2009, the U.S. incidence of *Campylobacter* infection was 13.0 per 100,000 population, with a rate in Tennessee of 8.2 cases per 100,000 population. Despite the lower overall incidence, Tennessee public health officials seek a better understanding of why and where this disease is occurring. Studies have been performed in Denmark, Norway and Canada to characterize spatial patterns of campylobacteriosis (Green et al., 2006; Jepsen et al., 2009; Jonsson et al., 2010). For example, Green et

al. (2006) detected the highest incidence in rural and agricultural regions with high densities of farm animals. The use of geographical information systems (GIS) and spatial statistics to identify high-risk regions serves to support and expand current knowledge on the mechanisms of transmission and geographical risk factors for this disease. Furthermore, geospatial tools provide epidemiologists and public health officials with a starting point from which to assess potential geographical associations and risk factors.

A major goal in the application of spatial statistics is to make meaningful inferences from available geographical data (Pascutto et al., 2000; Beale et al., 2008). For this reason, understanding the impact of spatial scale on the assessment of health outcomes is a topic of great importance. A number of studies exist which utilize cluster detection and mapping techniques to identify high-risk regions for infectious diseases (Lai et al., 2004; Odoi et al., 2004; Que and Tsui, 2008). Recent studies have also compared techniques for ecological analyses and employed a variety of geographical scales for understanding health outcomes and disparities (Krieger et al., 2002; Kitron et al., 2006; Kramer et al., 2010). However, few studies compare the differences in spatial patterns of disease at more than one level. Tian et al. (2010) explored the impact of geographical scale on breast cancer mortality among ethnic groups in Texas and found that the location of racial disparities changed depending on geographical level. This suggests that risk factors play different roles at different levels of aggregation. Odoi et al. (2003) compared giardiasis rates in Ontario, Canada, at two levels. Results showed spatial clustering of high rates at the smaller census sub-division scale, but not at the census division. Results of similar studies vary and there is no consensus on what geographical scale is most appropriate, precise or useful for surveillance data (Krieger et al., 2002; Elliott and Wartenberg, 2004; Oliver et al., 2005).

Despite inconsistency encountered in the literature, the choice of spatial scale may be crucial when investigating spatial patterns of disease. This choice is subject to the modifiable areal unit problem (MAUP) as well as the potential for ecological fallacy as described by Openshaw (1984). The MAUP applies to potential problems in both spatial scaling and zoning whereby conclusions and inferences may differ depending on the spatial level of analysis and type of areal divisions. From a public health standpoint, the utility of surveillance data is oftentimes limited by the level to which it can be aggregated and analyzed. Consistent analytical results over multiple geographical scales serve to vali-

date spatial statistical techniques as well as the reliability of the data used for a given study area. To our knowledge, no studies to date have been conducted to explore the direct impact of geographical scale on results of spatial cluster investigation using aggregated campylobacteriosis data.

Materials and methods

Study area and data sources

The study area encompassed the entire U.S. state of Tennessee, which has a land area of 66,333 km² and is the 36th largest out of 50 states. The estimated population in the year 2000 was 5,689,276 (approximately 63 people per km²) (U.S. Census Bureau, 2010). Campylobacteriosis surveillance data covering the period from 1 September 1991 to 31 December 2008 were provided by the Tennessee Department of Health. A case of campylobacteriosis was defined as culture-confirmed infection from a clinical specimen, typically stool or blood. Species and strain information were not reported. Participating laboratories are part of both a national and state-wide active surveillance system. This study received The University of Tennessee Institution Review Board approval, and all cases (n = 4,723) were de-identified to maintain anonymity and confidentiality.

Cartographical boundary files were obtained at the census tract (n = 1,261), zip code (n = 612) and county subdivision spatial levels (n = 462) from the U.S. Census Bureau Tiger Files (2000). These three spatial units are delineated as "statistical geographic entities" by the U.S. Census Bureau. Census tracts are the smallest units used in this study. They typically contain between 2,500 and 8,000 people and are relatively homogeneous with respect to population characteristics and socioeconomic conditions. The 5 digit zip code areas are categories used by the U.S. Postal Service to group mailing addresses. Cartographical boundaries, specifically zip codes boundaries and numbers are subject to change over time. Unlike census tracts and county subdivisions, which are both subunits of counties, zip codes may cross county boundaries and occasionally cross state lines. For instance, Tennessee contains one small zip code (42223, area 142.4 km²), located on the northern, central border which crosses the state boundary with Kentucky. All population denominator data, including the U.S. standard population used for risk standardization were obtained from the year 2000 U.S. Census.

Geocoding and data aggregation

The complete Tennessee dataset consisted of 4,723 campylobacteriosis cases. Initial data exploration for errors, repeat specimens, normality and outliers, as well as preparation and formatting for geocoding, was performed in SAS version 9.2 (2008). A total of 130 (2.7%) of the cases were identified as travel-related, 40 (0.8%) which lacked address information, and 24 (0.5%) repeat specimens were excluded during the cleaning and geocoding process. To allow for age and sex standardization, an additional four cases with missing sex information and 135 with missing age information were excluded from future analysis. A total of 967 (20.5%) cases were excluded from the dataset. From a geographical standpoint, these included 128 (2.7%) cases that contained information only at county level, 164 (3.5%) cases with only zip code level information, 228 cases (4.8%) containing only Post Office Box numbers and 244 cases (5.2%) with addresses that were not geocodable.

Geocoding was performed using both Googlemaps (2010) and Yahoo Maps Geocoder through BatchGeo (2010). An iterative process was undertaken to accurately match location data to the lowest level, or finest possible scale. This included manually matching address data to its exact latitude and longitude. The dataset utilized in the final analysis included 3,756 cases with 2,638 cases at rooftop accuracy and 1,118 cases at street level accuracy. An observation coded to the "rooftop" is precisely matched to its address on the ground. The "street level", also referred to as "range interpolated" accuracy, matched the case to an interpolated point on the road where the address would be located if the exact street number were unavailable. Each point was represented as a latitude and longitude coordinate and mapped in ArcGIS (ESRI, 2009). A point-in-polygon join was then used to merge the campylobacteriosis data to the census tract, zip code and county subdivision cartographical boundary files.

To assess for systematic differences in spatial distribution of cases retained in the dataset and those excluded, the observations in the unused portion of the dataset, with the exception of the 40 cases lacking location information, were mapped separately to the highest available accuracy level and visualized. No visual difference in spatial distribution was detected between the dataset used for analysis and the observations that were eliminated. By maintaining only the cases at the lowest, most accurate level of geocoding (rooftop and street), we sought to maximize the strength of the rela-

tionship between the three aggregate geographical levels and disease risk (Mazumdar et al., 2008).

Standardization

Campylobacteriosis typically demonstrates a bimodal age distribution (peaks occurring in children under 5 years of age and young adults), and it occurs more frequently in males (Allos and Taylor, 1998; Tauxe, 2001). In addition, gender-specific incidence has been shown to vary considerably among a wide variety of age groups (Gillespie et al., 2008). To account for these differences in risk, the population was grouped by sex and subdivided into six age categories: 0-4, 5-19, 20-39, 40-59, 60-74 ≥75 years. Age and sex standardization was then performed at all three geographical levels using STATA version 9.0 (2009) with adjusted risk estimates presented as number of cases per 100,000 population. The estimates were classified into five categories using Jenk's optimization method, a common statistical technique available within ArcMap (ESRI, 2009) which creates cut points based on inherent patterns in the data.

Spatial analysis

(a) Smoothing

Visualization of disease patterns can be distorted by spatial autocorrelation and excessively high variances often encountered in geographical areas with low disease counts and/or low background populations. To allow for better visualization, spatial empirical Bayesian (SEB) smoothing technique was implemented in GeoDa version 095i (Anselin, 2010) using counts obtained by the standardization procedure. The SEB smoothing was used to adjust for spatial autocorrelation and mitigate the small numbers problem (Odoi et al., 2003). The raw risk data were smoothed at each of the spatial levels, using first order Queen spatial weights. In keeping with Tobler's first law of geography (Tobler, 1970), Queen weighting gives greater influence to regions directly surrounding each areal unit of analysis. The SEB smoothing technique thereby maintains the integrity of the risk distribution without over-smoothing or creating a homogenous surface (Pascutto et al., 2000; Beale et al., 2008).

(b) Cluster detection

Cluster detection was performed using Kulldorff's spatial scan statistic, implemented in SaTScan version

8.2.1 (Kuldorff, 2010). The Poisson probability model was fit and estimates adjusted for sex and age in the analysis. Statistical significance was assessed at p-value <0.05 and performed using 999 Monte Carlo iterations. Scanning parameters included circular clusters of high-risk only, with no geographical overlap. There is a lack of comprehensive knowledge as to effect of the scanning window size on disease cluster results. For this study, the maximum scanning window size incorporated 3% of the total population at-risk and was chosen *a priori* with consideration to the low infectivity of the bacteria, geographical area and units of study. Window sizes of 5% and 10% were run post hoc to explore how cluster results vary by window size at the three geographical levels.

Results

Risk estimates and smoothing

County subdivision campylobacteriosis raw risk estimates ranged from 0 (n = 89) to 498 per 100,000 population (median = 55.0). Zip code area estimates ranged from 0 (n = 168) to 4,254 per 100,000 population (median = 49.8). Census tract risk estimates were highly variable, ranging from 0 (n = 241, median = 53.8) to 13,122 per 100,000 persons (raw risk maps not presented). Spatial empirical Bayesian smoothed maps (Figs. 1-3) demonstrate similar regions of elevated risk for the three spatial levels with the finest detail captured at the census tract level. From a broad perspective, these areas include the east Tennessee valley surrounding Knoxville, Cookeville and the region south of metropolitan Nashville-Davidson, in the central portion of the state. These regions are most clearly visualized at the county subdivision level (Fig. 1), due to a combination of large size polygons and tighter range in risk estimates. The highest risk class

category for both zip code and census tract spatial levels (Figs. 2 and 3) contained outliers after smoothing (as detected by box plots, not shown). These included one small urban zip code area (37902) located beside the Tennessee River within Knoxville city limits (Fig. 2c) and the two highest census tract level risks located in the regions of Center Point (smoothed risk 12,808.5 per 100,000) and Philadelphia (smoothed risk 6196.7 per 100,000) as indicated in dark purple (Fig. 3a-d).

Cluster detection

According to cluster detection results, the high risk areas (P <0.05) surround Knoxville, Maynardville and Jefferson City, south of metropolitan Nashville-Davidson, Cookeville, Murfreesboro, Shelbyville, Franklin and Oak Ridge (Fig. 4 a-c and Table 1). Regions east and west of the city of Knoxville were detected as clusters at all three spatial scales. Shelbyville, Kingsport and Lawrenceburg were identified as unique, singular, census tract clusters (Table 1) with very high relative risks (12.1, 5.8 and 9.0, respectively). The Kingsport and Lawrenceburg areas were only identified as high-risk clusters at the census tract level, with Kingsport being too small for visual identification on the state map. In comparison, Kingsport ranked 29th and Lawrenceburg ranked 19th highest out of the 1,261 tracts from the census tract unsmoothed, standardized risk results. The other single tract cluster is represented by the city of Shelbyville (ranked 12th highest in risk). The Shelbyville area is included as part of a larger cluster identified in zip code and county subdivision analyses.

An overlay map of the clusters (Fig. 5 a-c) shows that the high-risk clusters occur at consistently similar locations around Cookeville, east Knoxville and south Murfreesboro at all three spatial levels (labels excluded to aid visualization). The Franklin area, seen at the
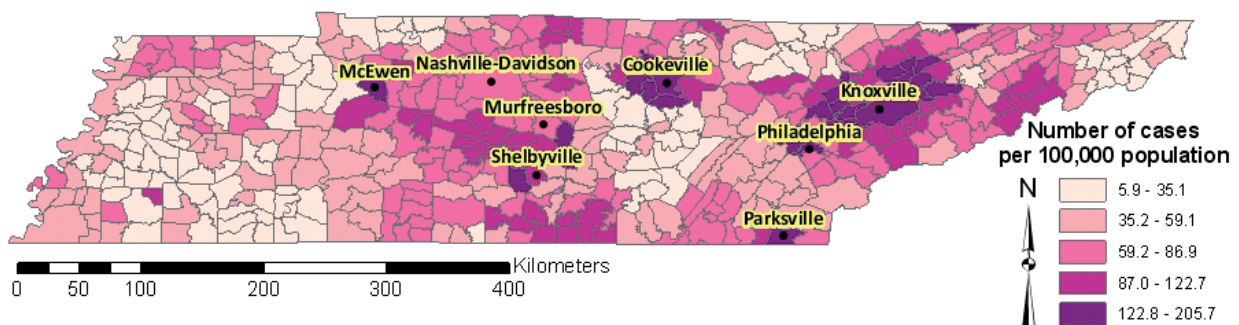
Data Sources: U.S. Census Bureau & Tennessee Department of Health

Fig. 1. Spatial empirical Bayesian smoothed risk of campylobacteriosis in Tennessee during the study period 1 September 1991 to 31 December 2008 at the county subdivision spatial level. The risks are standardized by age and sex.
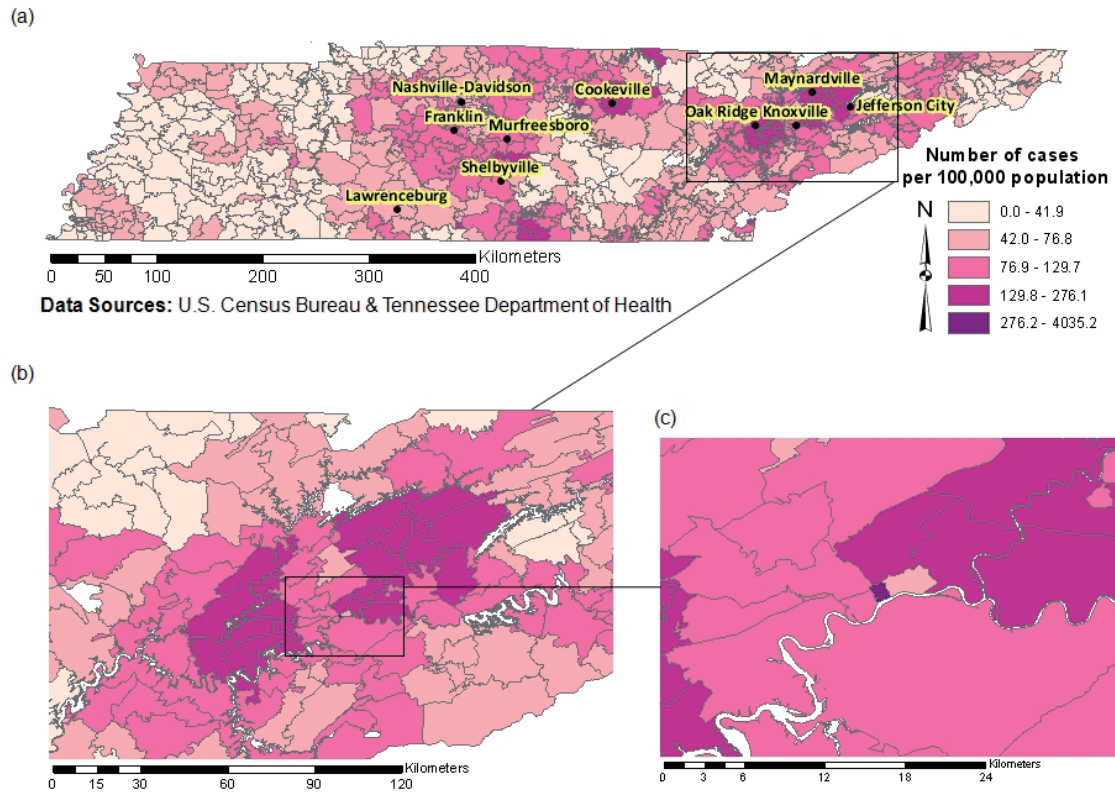
Fig. 2. (a) Spatial empirical Bayesian smoothed risk of campylobacteriosis in Tennessee at the zip code spatial level. (b) Primary insert incorporates the area surrounding Knoxville. (c) Secondary insert highlights the highest risk region of central Knoxville (dark purple). The risks are standardized by age and gender.
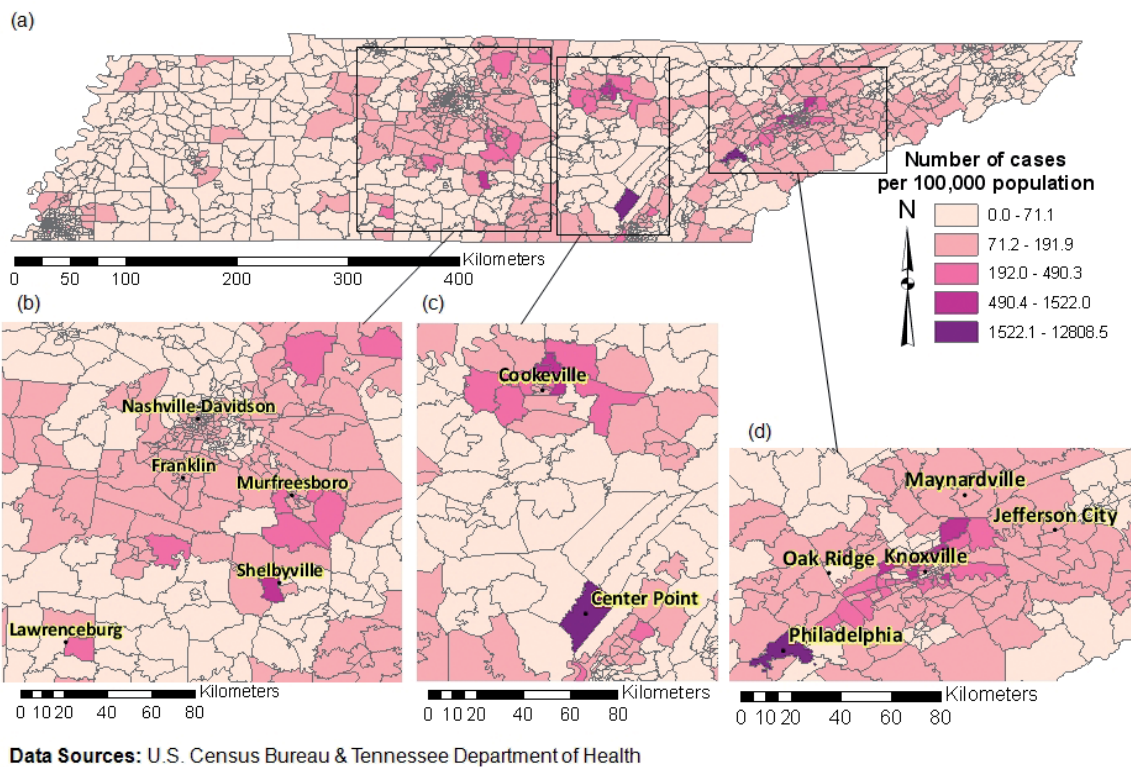

Fig. 3. (a) Spatial empirical Bayesian smoothed risk of campylobacteriosis in Tennessee at the census tract spatial level. The three regional inserts (b-d) highlight important high risk areas at this scale. The risks are standardized by age and gender.

(a)

(b)

(c)

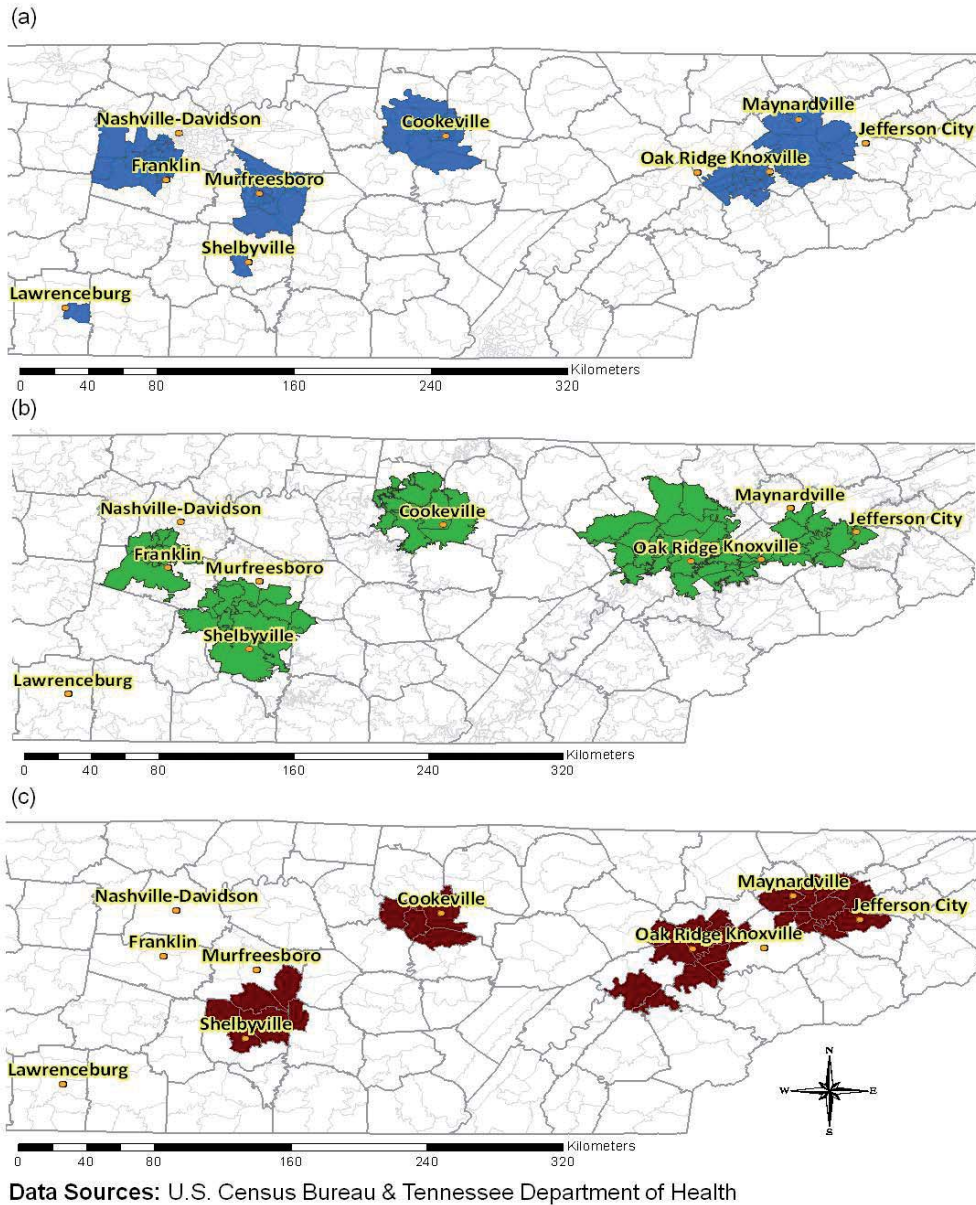**Data Sources:** U.S. Census Bureau & Tennessee Department of Health

Fig. 4. Significant high-risk clusters using 3% scanning window at three levels: (a) county subdivision, (b) zip code, (c) census tract.

top most, western cluster contains a comparable overlay at the zip code and census tract levels, but not county subdivision. Other differences can be visualized in the easternmost cluster (cluster 3, Table 1). At the zip code level this cluster encompasses a larger region of high risk, north and west of Oak Ridge (Fig. 5). Furthermore, county subdivision analysis uniquely detected the Philadelphia area (cluster 5, Table 1).

Alternate scanning window analysis

The larger scanning window sizes of 5% and 10% resulted in clusters covering similar regions, with some variation, at all three levels. As an example, consecu-

tively larger cluster sizes (incorporation of greater number of areal units) were found at the county subdivision level as the window size was increased (Fig. 6 a-c). Similar overall expansion of cluster size was found at the census tract and zip code levels (maps not shown). Regardless of the scan window size, the Lawrenceburg and Kingsport regions were not identified as clusters in the county subdivision analysis.

Discussion

The results of both the smoothing and spatial scan analyses show the general regions surrounding Knoxville and Cookeville, and south of metropolitan

Table 1. High risk campylobacteriosis cluster profiles at census tract, zip code and county subdivision spatial levels.

	Cluster (P-value)	General Region	Observed[1]	Expected[2]	RR[3]	Count[4]	Population
	1*(P = 0.001)	Knoxville	289	106.5	2.86	32	161,266
	2 (P = 0.001)	Cookeville	97	28.3	3.43	14	42,008
	3 (P = 0.001)	Maynardville, Jefferson City	185	80.6	2.36	37	121,862
	4 (P = 0.001)	Murfreesboro	122	46.0	2.71	18	67,564
Census tract	5 (P = 0.001)	Shelbyville	16	1.3	12.06	1	2,010
	6 (P = 0.001)	Nashville-Davidson, Franklin	176	106.2	1.69	33	167,671
	7 (P = 0.001)	KingsPort	16	2.8	5.82	1	4,181
	8 (P = 0.003)	Lawrenceburg	10	1.1	9.00	1	1,775
	1*(P = 0.001)	Knoxville	267	110.8	2.52	8	12,627
	2 (P = 0.001)	East Knoxville, Jefferson City	157	74.6	2.15	12	114,482
	3 (P = 0.001)	Oak Ridge	179	96.6	1.90	16	149,081
Zip code	4 (P = 0.001)	Cookeville	97	45.6	2.16	7	70,054
	5 (P = 0.001)	Nashville-Davidson, Franklin	182	110.0	1.69	6	168,288
	6 (P = 0.001)	Shelbyville, Murfreesboro	98	56.3	1.76	11	83,231
	1* (P = 0.001)	West Knoxville	207	98.2	2.17	11	151,917
	2 (P = 0.001)	Maynardville	104	43.5	2.43	10	66,116
County sub-division	3 (P = 0.001)	Cookeville	89	38.9	2.32	4	59,547
	4 (P = 0.002)	Shelbyville, Murfreesboro	61	30.3	2.03	7	11,215
	5 (P = 0.009)	Philadelphia, Barnard	19	5.5	3.47	2	8,433

[1] The observed number of cases within the cluster; [2] the expected number of cases as calculated by the spatial scan algorithm; [3] the relative risk of campylobacteriosis for the cluster; [4] denotes the number of areal units included in the cluster; *denotes the primary cluster detected in the analysis.

Nashville-Davidson to be at high risk for campylobacteriosis in Tennessee, at all three spatial levels. Previous research suggests that the risk of acquiring campylobacteriosis may be higher in different areas due to the presence of different risk factors (Green et al., 2006; Jepsen et al., 2009; Jonsson et al., 2010). For example, in metropolitan areas such as Knoxville, one might investigate centralized sources of contaminated poultry products. In rural areas exposure to domestic farm animals and local environmental sources might be of greater importance (Odoi et al., 2004; Green et al., 2006). Addressing the impact of socioeconomic factors and quantifying potential health inequalities between advantaged and disadvantaged groups could be applied to high risk areas across the state. The specific reasons for the geographic differences found in this study are unknown and should be investigated further.

The fact that all three levels share comparable geographic clustering is a unique strength of this study. Other studies performing spatial analyses at more than one spatial scale demonstrate differences in patterns of health outcomes and associations across spatial scales (Openshaw, 1984; Krieger et al., 2002; Tian et al., 2010). This may be a result of the MAUP, data errors and missing data, as well as variation in either the disease data or the population and ecological level data (Krieger et al., 2002; Elliott and Wartenberg, 2004; Beale et al., 2008). While spatial variation in risk exists in some regions, these results highlight the overall similarity in spatial patterns detected between geographic levels using clustering and smoothing techniques simultaneously. Our study demonstrates that consistent results across spatial scales occur despite the inherent limitations encountered in ecological analyses. While the results are specific to campylobacteriosis in Tennessee, these findings may lend credence to studies that utilize disease surveillance data at only one level (Green et al., 2006; Jepsen et al., 2009; Sugumaran et al., 2009).

The lower level (census tract) analysis produces the finest detail and is able to capture small, yet potentially important areas of high risk. This conclusion is supported by previous studies which found small area studies to be the most useful for measuring health disparities and reducing ecological bias (Reading and Openshaw, 1993; Beale et al., 2008; Tian et al., 2010). Where possible, analyses should be conducted at more than one level to assess the full spatial picture. This is particularly important for validating results from finer
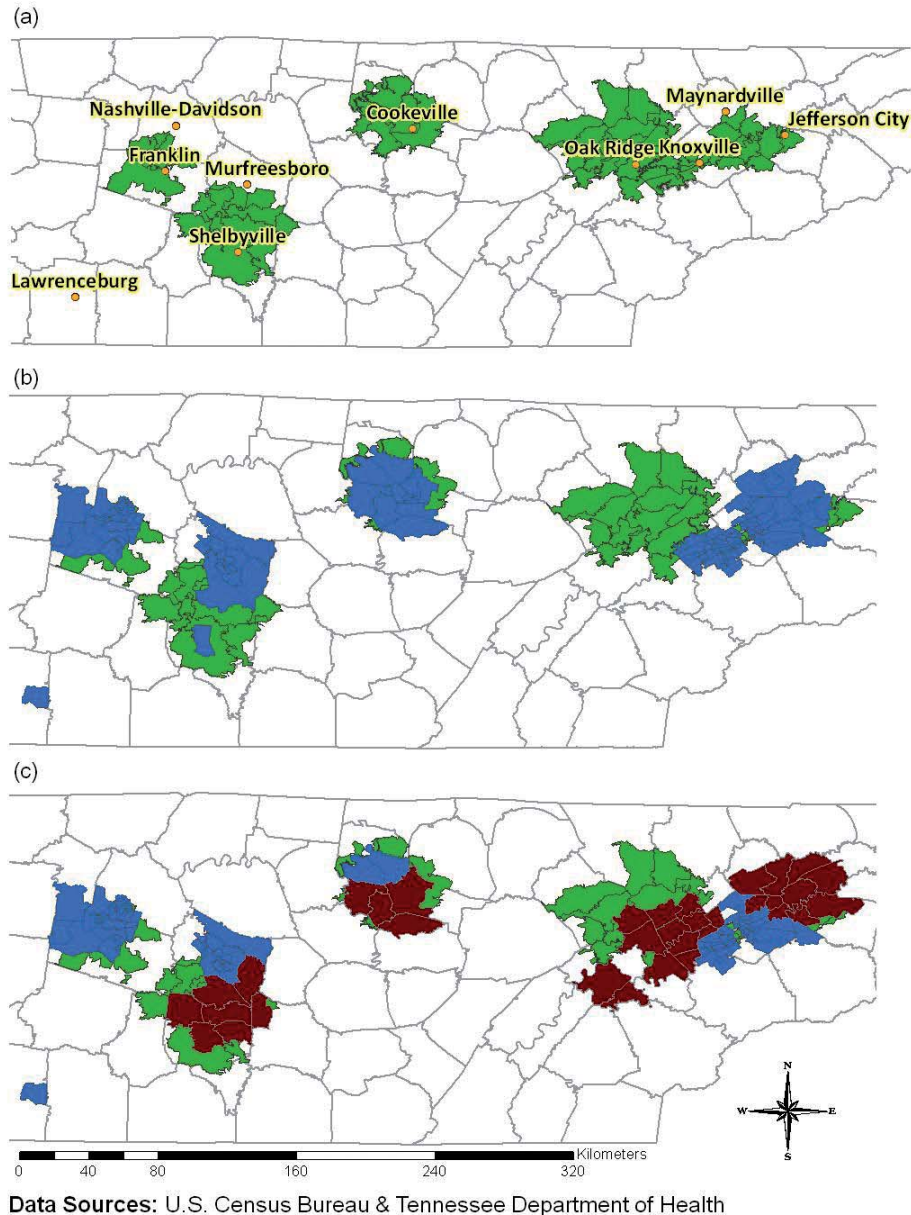
Fig. 5. Magnified view of Tennessee demonstrates simultaneous overlay of clusters using 3% scanning window: (a) includes only the zip code (green), (b) contains the zip code and census tract (blue) and (c) shows all three (from bottom to top): zip code, census tract and county subdivision (brown). Note: order of overlay was chosen for improved visualization.

scale levels where positional inaccuracy can be more pronounced and lead to decreased precision (Oliver et al., 2005). Furthermore, socioeconomic, demographic and environmental data are often unavailable at finer scales, and the process of geocoding can be costly and time consuming. When census tract analyses are either undesirable or not feasible, the zip code or county subdivision aggregations may provide sufficient information for cluster detection and prevention and control of campylobacteriosis in Tennessee.

In a recent comparison of cluster detection software methods, SaTScan was found to be a fast, user-friend-

ly and well-developed for cluster detection (Robertson and Nelson, 2010). The spatial scan statistic allows the window to vary during the scan process. This methodology decreases the chance of pre-selection bias as clusters may be detected without prior knowledge of geographical extent (Kulldorff, 1997). The fact that size, shape and number of clusters changed in our study when the scan window size was enlarged is a complication seen with use of this method. Understanding the implications of scanning window size on clustering is beyond the scope of this paper. However, recent studies focus on result variability
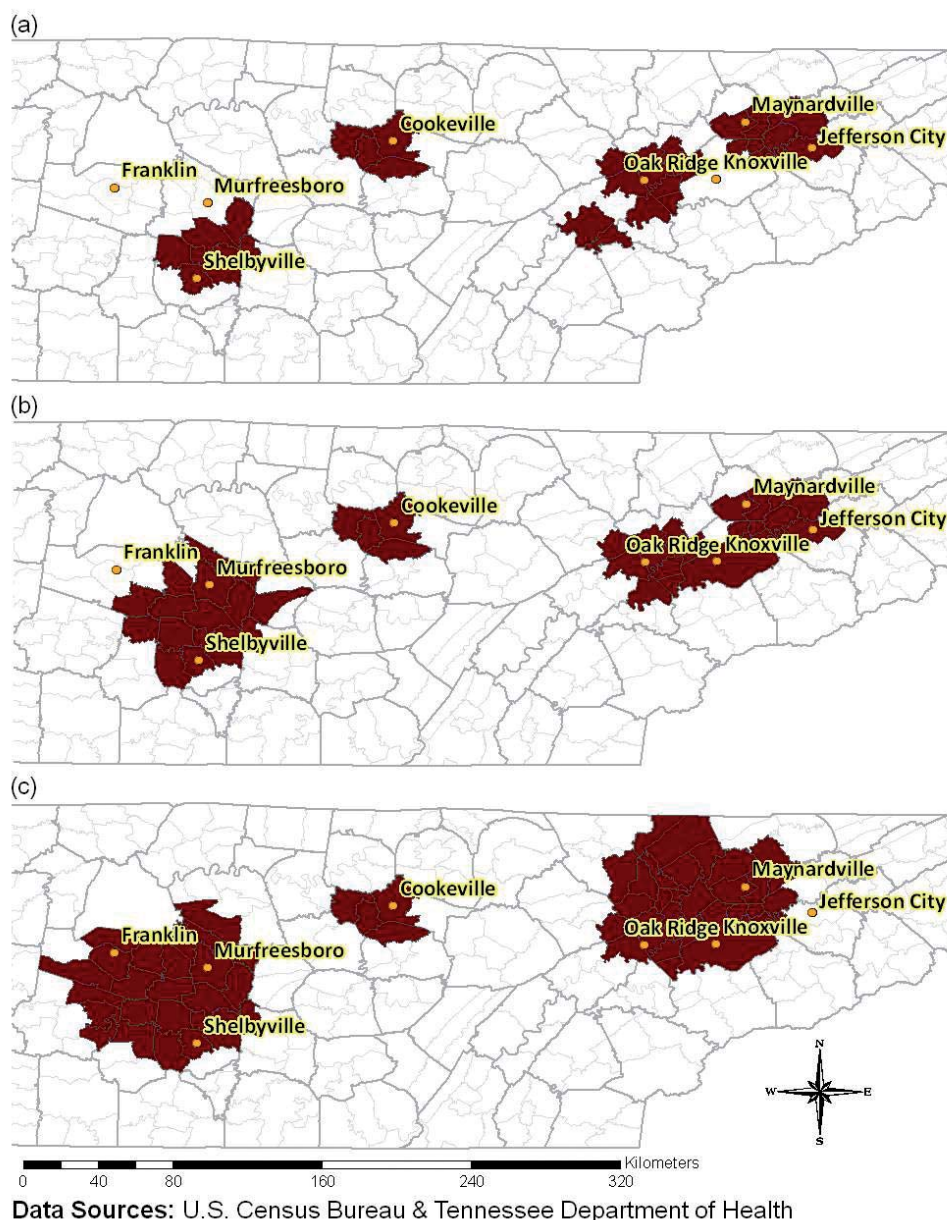
Fig. 6. Significant (P <0.05) high-risk cluster results for county subdivision spatial level using (a) 3% scanning window, (b) 5% scanning window, (c) 10% scanning window.

dependent on scan window size (Chen et al., 2008; Sugumaran et al., 2009). Further exploration into both the theoretical and practical implications of scan window choice would add to the growing body of knowledge on the spatial scan statistic. Consecutive analyses can lead to improved understanding of *Campylobacter* distribution over time and space (Jonsson et al., 2010). In our study, we interpret the SEB smoothed risk distribution in conjunction with cluster analysis at three levels to offer a more comprehensive visual and statistical approach to understanding campylobacteriosis disease patterns.

Cluster detection results suggest that environmental and/or socioeconomic factors contribute to disease transmission. A model-based approach may be applied to campylobacteriosis surveillance data as a means of assessing potential associations between disease risk and epidemiological factors (Pascutto et al., 2000; Green et al., 2006). Furthermore, mapping risk and cluster analysis results on all three levels with similar findings serves to statistically validate the location of high-risk regions in Tennessee. Previous ecological studies have shown that associations between risk factors and disease may be dependent on the level of

aggregation (Odoi et al., 2004; Tian et al., 2010). Modeling at more than one spatial level, therefore, serves to enhance current knowledge of disease patterns and provide a more accurate link between important disease predictors and a constantly changing environment. Temporal variations in campylobacteriosis risk may also be important in understanding transmission routes and changing distribution patterns (Jepsen et al., 2009; Jonsson et al., 2010). In this study, time components were not examined, as partitioning the surveillance data into temporal units would result in insufficient data for meaningful analysis.

A number of factors may play a role in the accuracy and usefulness of this study. Many ill persons do not seek care, have a stool culture performed, or have the etiology identified. In addition, data reporting may vary and the impact of using census year 2000 as denominator data for cases spanning 17 years is not known. Underreporting is expected for campylobacteriosis as the disease is typically self limiting and very rarely fatal. These data include cases ascertained by active surveillance, and are thought to accurately represent the burden of culture-confirmed campylobacteriosis (Scallan et al., 2011). Approximately 20% of the case data for this study were eliminated due to missing information or data address inaccuracy and the effect of this loss is difficult to quantify. The spatial distribution of cases with missing data was similar to those whose data were complete, implying that the missing data were randomly distributed. Misclassification of cases may also occur and may be a result of data inaccuracies, human migration and changes in geographical boundaries (Oliver et al., 2005; Beale et al., 2008). Furthermore, a large number of cases arising from a common source may result in a regional elevation in disease risk that distorts the location of clustering. Improved identification and documentation of cases due to an outbreak would help to eliminate this potential source of bias as well as add to our knowledge of the characteristics of the population at risk. Recently, FoodNet began collecting data on cases being outbreak-associated. This information was not available for all years analyzed. These inherent limitations add uncertainty and potential bias to spatial analysis research.

Patterns of campylobacteriosis may change over time in response to changing human and animal demographics as well as shifts in climatic factors and land use. Sequential analyses are warranted to improve our understanding of the disease. Our research is intended to enhance current knowledge of the human campylobacteriosis distribution and risk, and to create a foundation upon which ecological associations, including socioeconomic and environmental factors, can be superimposed on clusters to identify risk factors. Technological advances in statistical and mapping software and the growing availability of surveillance data justify continual review of spatial methods, the levels at which these methods are applied, and their applicability to public health.

In conclusion, the findings of this study demonstrate an overlap in spatial clustering of campylobacteriosis across three geographic scales in Tennessee. Some variation in the size and shape of clusters was present. However, the overall disease patterns were similar, leading to improved confidence in comparisons between levels. This finding is especially important in situations where address data are unavailable, making finer scale analyses impossible. Identification and superimposition of regional-specific risk factors would help generate hypotheses for the spatial differences in campylobacteriosis risk identified in these analyses. Visual comparison of smoothed risk estimates and cluster detection at multiple spatial levels has the potential to help public health officials effectively identify geographical, socioeconomic and environmental factors which may play an important role in the occurrence of campylobacteriosis. Subsequently, this knowledge could be used to create a framework whereby future disease control strategies become more geographically precise and economically efficient.

Acknowledgements

References

Allos BM, Taylor DN, 1998. *Campylobacter* infections. Bacterial infections of humans, epidemiology and control, Plenum Medical Book Company, 169-190.

Altekruse SF, Swerdlow DL, 2002. *Campylobacter jejuni* and related organisms. Foodborne Diseases, Academic Press, 103-112.

Anselin L, 2010. GeoDa0.9.5-i . Available at: http://geodacenter.asu.edu/software/downloads. Arizona State University, USA.

Batchgeo, 2010. Available at: http://www.batchgeo.com/(accessed on May 2010).

Beale L, Abellan JJ, Hodgson S, Jarup L, 2008. Methodologic issues and approaches to spatial epidemiology. Environ Health Perspect 116, 1105-1110.

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.

Elliott P, Wartenberg D, 2004. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect 112, 998-1006.

ESRI, 2009. Arcview GIS version 9.2. Environmental Systems Research Institute, Inc., Redlands, California, USA.

Frost J, 2001. Current epidemiological issues in human campylobacteriosis. Symp Ser Soc Appl Microbiol, 85S-95S.

Google Maps, 2010. Available at: http://maps.google.com/ (accessed on May 2010).

Gillespie IA, 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, 1717-1725.

Green C, Krause D, Wylie J, 2006. Spatial analysis of *Campylobacter* infection in the Canadian province of Manitoba. Int J Health Geogr 5, 2.

Hearnden M, Skelly C, Eyles R, Weinstein P, 2003. The regionality of campylobacteriosis seasonality in New Zealand. Int J Environ Health Res 13, 337-348.

Humphrey T, O'Brien S, Madsen M, 2007. Campylobacters as zoonotic pathogens: a food production perspective. Int J Food Microbiol 117, 237-257.

Jepsen MR, Simonsen J, Ethelberg S, 2009. Spatio-temporal cluster analysis of the incidence of *Campylobacter* cases and patients with general diarrhea in a Danish county, 1995-2004. Int J Health Geogr 8, 11.

Jonsson ME, Heier BT, Norstrom M, Hofshagen M, 2010. Analysis of simultaneous space-time clusters of *Campylobacter* spp. in humans and in broiler flocks using a multiple dataset approach. Int J Health Geogr 9, 48.

Kitron U, Clennon JA, Cecere MC, Gurtler RE, King CH, Vazquez-Prokopec G, 2006. Upscale or downscale: applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya. Geospat Health 1, 49-58.

Kovats R, Edwards S, Charron D, Cowden J, D'Souza R, Ebi K, 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, 207-214.

Kramer MR, Cooper HL, Drews-Botsch CD, Waller LA, Hogue CR, 2010. Do measures matter? Comparing surface-density-derived and census-tract-derived measures of racial residential segregation. Int J Health Geogr 9, 29.

Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R, 2002. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. Am J Epidemiol 156, 471-482.

Kulldorff M, 1997. A spatial scan statistic. Commun Stat Theory Methods 26, 1481-1496.

Kulldorff M and Information Management Services, 2010. SaTScanTM v8.0: Software for the spatial and space-time scan statistics. Available at: http://www.satscan.org/download.html (accessed on April 2010)

Lai PC, Wong CM, Hedley AJ, Lo SV, Leung PY, Kong J, Leung GM, 2004. Understanding the spatial clustering of severe acute respiratory syndrome (SARS) in Hong Kong. Environ Health Perspect 112, 1550-1556.

Mazumdar S, Rushton G, Smith BJ, Zimmerman DL, Donham KJ, 2008. Geocoding accuracy and the recovery of relationships between environmental exposures and health. Int J Health Geogr 7, 13.

Nylen G, Dunstan F, Palmer S, Andersson Y, Bager F, Cowden J, Feierl G, Galloway Y, Kapperud G, Megraud F, Molbak K, Petersen L, Ruutu P, 2002. The seasonal distribution of campylobacter infection in nine European countries and New Zealand. Epidemiol Infect 128, 383-390.

Odoi A, Martin SW, Michel P, Holt J, Middleton D, Wilson J, 2003. Geographical and temporal distribution of human giardiasis in Ontario, Canada. Int J Health Geogr 2, 5.

Odoi A, Martin SW, Michel P, Middleton D, Holt J, Wilson J, 2004. Investigation of clusters of giardiasis using GIS and a spatial scan statistic. Int J Health Geogr 3, 11.

Oliver MN, Matthews KA, Siadaty M, Hauck FR, Pickle LW, 2005. Geographic bias related to geocoding in epidemiologic studies. Int J Health Geogr 4, 29.

Openshaw S, 1984. Ecological fallacies and the analysis of areal census data. Environ Plan A 16, 17-31.

Pascutto C, Wakefield JC, Best NG, Richardson S, Bernardinelli L, Staines A, Elliott P, 2000. Statistical issues in the analysis of disease mapping data. Stat Med 19, 2493-2519.

Que J, Tsui FC, 2008. A multi-level spatial clustering algorithm for detection of disease outbreaks. AMIA Annu Symp Proc, 611-615.

Reading R, Openshaw S, 1993. Do inaccuracies in small area deprivation analyses matter? J Epidemiol Community Health 47, 238-241.

Robertson C, Nelson TA, 2010. Review of software for space-time disease surveillance. Int J Health Geogr 9, 16.

SAS Institute Inc, 2008. Statistical analysis systems (SAS) version 9.2 . Cary, North Carolina, USA.

Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM, 2011. Foodborne illness acquired in the United States—major pathogens. Emerg Infect Dis 17, 7-15.

Snelling W, Matsuda M, Moore J, Dooley J, 2005. *Campylobacter jejuni*. Lett Appl Microbiol 41, 297-302.

STATA Corporation LP, 2009. Stata version 10.0. Available at: http://www.stata.com/ College Station, Texas, USA.

Sugumaran R, Larson SR, Degroote JP, 2009. Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. Int J Health Geogr 8, 43.

Tauxe RV, 2001. Major risk factors for human campylobacteriosis-an overview. WHO Consultation on the Increasing Incidence of Human Campylobacteriosis. The Increasing Incidence of Human Campylobacteriosis: Report and Proceedings of a WHO Consultation of Experts, Copenhagen, Denmark 21-25 November 2000. World Health Organization, 65-66.

Tian N, Goovaerts P, Zhan FB, Wilson JG, 2010. Identification of racial disparities in breast cancer mortality: does scale matter? Int J Health Geogr 9, 35.

Tobler WR, 1970. A computer movie simulating urban growth in the Detroit region. Econom Geogr 46, 234-240.

U.S. Census Bureau, 2010. Available at: http://factfinder.census.gov /home/saff/main.html?_lang=en. (accessed on June 2010).

U.S. Census TIGER Line Data, 2000. Available at: http://www.census. gov/geo/www/tiger/tiger2k/tgr2000.html. (accessed on May 2010).

Vally H, Hall G, Scallan E, Kirk MD, Angulo FJ, 2009. Higher rate of culture-confirmed *Campylobacter* infections in Australia than in the USA: is this due to differences in health-care-seeking behaviour or stool culture frequency? Epidemiol Infect 1, 1-8.

Weisent J, Seaver W, Odoi A, Rohrbach B, 2010. Comparison of three time-series models for predicting campylobacteriosis risk. Epidemiol Infect 138, 898-906.