An evaluation of SaTScan for the prospective detection of space-time *Campylobacter* clusters in the North East of England

G. J. HUGHES[1,2]* AND R. GORTON[1]

[1] *Health Protection Agency North East, Citygate, Gallowgate, Newcastle upon Tyne, UK*
[2] *Field Epidemiology Training Programme, Health Protection Agency, Colindale, London, UK*

SUMMARY

Campylobacteriosis is the commonest cause of bacterial enteritis in England yet the epidemiology of apparently sporadic cases is not well understood. Here we evaluated the feasibility of applying a space-time cluster detection method to routine laboratory surveillance data in the North East of England by simulating prospective weekly space-time cluster detection using SaTScan as if it had been performed for 2008–2011. From the 209 simulated weekly runs using a circular window, 20 distinct clusters were found which contained a median of 30 cases (interquartile range 15–66) from a median population of ∼134 000 persons. This corresponds to detection of a new cluster every 10 weeks. We found significant differences in age, sex and deprivation score distributions between areas within clusters compared to those without. The results of this study suggest that space-time detection of *Campylobacter* clusters could be used to find groups of cases amenable to epidemiological investigation.

Key words: *Campylobacter*, epidemiology, outbreaks, spatial modelling, surveillance.

INTRODUCTION

Campylobacteriosis is the commonest cause of bacterial enteritis in England with almost 63 000 laboratory-confirmed cases in 2010 [1]. Globally, campylobacteriosis is a seasonal infection [2] and in England there is a major annual peak in June/July [3]. Most cases are assumed to be sporadic infections which lack epidemiological linkage [2]. In England, international travel and consumption of chicken in restaurants are strongly linked to sporadic infections [4] which is similar to the situation reported elsewhere [5]. Worldwide, the consumption of chicken may

account for 40–70 % of human infections [5] while for cases reported in England, 17 % are associated with foreign travel [6].

The majority of analytical studies of human *Campylobacter* infection have focused on food/drink consumption and contact with companion animals and livestock [5]. Avian species are thought to be the true natural hosts of *Campylobacter* spp. but there is considerable evidence for infection of cattle and sheep and genetic evidence of transmission to humans [7]. Studies have demonstrated a number of environmental risks for campylobacteriosis such as environmental contamination from *Campylobacter*-infected broiler farms [8], high ruminant density [9] and frequent contact with chickens [10]. *Campylobacter* can survive in faeces and the environment [11, 12] and rainfall has been associated with contamination of surrounding

* Author for correspondence: Dr G. J. Hughes, Health Protection Agency North East, Citygate, Gallowgate, Newcastle upon Tyne NE1 4WH, UK.
(Email: gareth.hughes@hpa.org.uk)

areas [13] where *Campylobacter* survival can occur for long periods of time [14]. In England there is conflicting evidence as to whether the number of infections in poultry farms shows substantial fluctuation [15] but it seems unlikely that the seasonality in the number of human campylobacteriosis cases is a direct consequence of a coincident trend in the number of infections of birds on poultry farms [3].

A relatively small number of campylobacteriosis outbreaks are reported each year in England [16]. Those identified tend to be foodborne outbreaks associated with the catering industry (e.g. [17]) where, as with other detected foodborne outbreaks, information provided by the general public is foremost for recognition [18]. Given the high number of laboratory-confirmed cases, it is possible that community outbreaks occur which are not detected [19]. Any such outbreaks would represent missed opportunities for improving understanding of the epidemiology of *Campylobacter* infection and for identification of measures to protect public health. Means to identify possible outbreaks which allow for targeted investigation will therefore be a valuable addition to existing surveillance strategies.

Space-time cluster detection provides a means by which clusters of cases (based on a location such as place of residence) can be detected early and in some situations detected where they would not have been detected by purely spatial or temporal methods. SaTScan [20] is an easy to use software program which meets the requirements for space-time disease surveillance without the additional need for substantial technical expertise [21] and can be readily applied to infectious disease surveillance. The spatial scan statistic used by SaTScan assesses whether the number of cases within a space-time cylinder (where the base of the cylinder is a circle or ellipse covering an area of the study region) are greater than expected given the number of cases which occur within the rest of the study region over the same time period [22, 23]. As such, these algorithms do not require adjustment for seasonal trends, inclusion of historical baseline data or prior definition of strictly bounded areas where clusters are to be detected. Clusters up to a maximum specified size can be detected anywhere within the study region in a single operation.

For each SaTScan run, the maximum-likelihood cluster is detected and the significance of this cluster compared to the rest of the study region using randomization testing. Secondary clusters with reduced likelihood are also reported with a corresponding significance determined in the same manner as for the most likely cluster [22]. Prospective space-time cluster detection models can be applied within SaTScan which will scan for clusters which are 'alive' during the most recent time period [24] and can be used by public health departments and epidemiology units as part of routine surveillance (e.g. [25]).

This study was undertaken to assess the value of such an approach for the detection of space-time clusters of campylobacteriosis for investigation in the North East of England. This is dependent first on whether the approach can detect clusters which it is feasible to investigate and second whether on investigation those clusters are found to be associated with common links or risk factors which shed light on transmission routes and/or provide opportunities for the protection of public health. To this end, the purpose of the current study was to simulate prospective space-time cluster detection for laboratory-confirmed cases of campylobacteriosis in the North East of England (a population of ∼2·5 million people) as if it had been used weekly over a 4-year period (2008–2011) in order to determine whether clusters would be found and to assess the feasibility of performing epidemiological and microbiological investigation of clustered cases.

METHODS

Surveillance data

Data for laboratory-confirmed cases of *Campylobacter* spp. infection resident in the North East of England were obtained from the Health Protection Agency North East regional surveillance system which collects data on selected laboratory and clinician-notified cases of infectious disease of public health interest. The best available onset date (herein, referred to simply as onset date) for each case is determined from a hierarchy of reported onset date, specimen date, result date and referral date. Given the need for correct and complete geographical placement of cases, all residential postcodes for cases were checked for validity prior to analysis. UK postcodes correspond to a single address, a street or part of a street and have associated geographical coordinates. Age and sex were also obtained for each case. Exposure data for potential risk factors is not routinely collected for cases of *Campylobacter*.

Space-time cluster detection

Cluster detection was performed using SaTScan v. 9.1.1 [20] with a discrete space-time Poisson model [22] for prospective cluster detection (where clusters must contain cases occurring within the most recent included time window). Under the null hypothesis of the Poisson model the number of expected cases is proportional to the specified population size, therefore accounting for variation in population density. Cluster detection used UK lower super output area (LSOA) of residence as the geographical unit; LSOAs are small statistical boundary areas of ∼1500 people. Office for National Statistics (ONS) all ages population estimates for the previous year (to simulate the real-time situation) were used as population sizes for the Poisson model and the easting and northing of the ONS LSOA population centroid from April 2005 (the most recently available) as Cartesian coordinates. The residential address for cases was linked to LSOA using the ONS Postcode Directory, 2011 Edition. All ONS data used is freely available (www.ons.gov.uk).

We simulated weekly space-time cluster detection runs for 209 separate weeks corresponding to each Monday from 1 January 2008 to 2 January 2012. For each run, data was censored to include only laboratory-confirmed cases which had been reported to the surveillance system before or during the previous calendar week. Onset date was used as a proxy for exposure date. Detection was performed using SaTScan models with either a 1-day or 7-day temporal window (separate analyses), a maximum population size of 50% of the region, a maximum cluster duration of 28 days, a circular or elliptical window (separate analyses) and a restriction of secondary clusters to those with a different centre to the most likely cluster.

Potential epidemiological clusters

The geographical area covered by cluster windows can overlap with other cluster windows both from the same week and from subsequent weekly runs. Hence the total number of significant clusters detected may be greater than the number of epidemiological clusters that would require investigation. In order to determine this number, clusters for analysis were restricted to either (1) a single geographical window with no geographically overlapping windows in the same and contiguous weeks, or (2) a set of closely overlapping geographical windows from one or more weeks. Multiple clusters detected in the same week but with multiple foci, and which were not all included in a single larger window, were treated as multiple epidemiological clusters. In considering whether or not significant clusters represented those of distinct epidemiological interest, we did not consider the duration of each cluster (i.e. the start date of the cluster was not taken into account). For each unique cluster defined in this way we recorded (1) the geographically largest initial window (i.e. detected on the first week of the cluster set), (2) the geographically largest window (detected on any week the cluster was detected) and (3) the highest probability window (i.e. that with the lowest associated P value).

In order to remove any potential bias in cluster detection due to the presence of multiple cases within a single household, SaTScan runs which detected a significant cluster were repeated after removal of excess household cases both from within the cluster and from the rest of the region for the time period during which a cluster was detected. Multiple household cases may represent instances of person-to-person transmission and could wrongly inflate the relevant surveillance signal for cases which are a result of a putative primary exposure. Household cases were defined as cases with the same postcode and house/flat number with a date of onset within the duration of a suggested cluster. When removing excess household cases the case with the earliest date of onset was retained; where onset dates were identical one or more cases were removed at random.

Comparison of clustered and non-clustered cases

A comparison between cases within clusters and cases reported elsewhere in the region during the same time period was performed using the highest probability window as the basis for each cluster. We compared age, sex and the deprivation scores of LSOAs.

Age distribution bootstrapping

In order to determine whether or not the age distribution of clustered cases was different to the age distribution of cases in the rest of the entire study region during the same time period, we simulated 10 000 age distributions of equal size for each cluster by randomly drawing from all cases within the region during the time period of the cluster. This approach allowed an assessment of differences in the observed number of cases from expected numbers for defined age groups. Age distributions were analysed according to

epidemiologically meaningful age groups derived from the incidence rate of cases by age for the time period of the study using the median mid-year population size from those years where data was available (2008–2010). For each cluster a significant departure from expectation for any age group was defined as an observed count in any age group being less than or greater than the interval covering 95% of the simulated counts.

Indices of multiple deprivation and rural/urban classification

Indices of multiple deprivation (IMD) 2010 scores [26] for all LSOAs whose centroid was contained within each cluster circle (irrespective of whether or not cases were reported from that LSOA) were compared to all other LSOAs within the region using an exact permutation test. Rural and urban area classification (RUAC) [27] for each cluster was compared to that of the entire North East region.

Statistical analysis

All data manipulation and statistical analysis was performed using R v. 2.10.1 [28].

RESULTS

Surveillance data

Of the 12 364 cases of campylobacteriosis that were reported to HPA North East during the study period (2008–2011), eight (0·06%) had missing home postcodes and were excluded from the analysis. Onset date was based on reported onset date for 24% of cases ($n = 2987$), specimen date for 73% ($n = 9035$), result date for 1% ($n = 147$) and referral date for 2% ($n = 195$). Cases showed a clear seasonal pattern with a minor peak in spring (February–March) and a major peak in summer (June–July) peaks (Fig. 1).

Incidence rates and age groups

Cases were divided into age groups (0–4, 5–19, 20–44, 45–64, ⩾65 years) based on the incidence rate (Fig. 2). A total of 25 cases were excluded from the analysis as they had either unknown sex ($n = 2$) or unknown age ($n = 23$). No significant differences were found between the age distribution of cases for males and females using quinary age groups or the age groups used in this study (results not shown).

Space-time cluster detection

From the 209 model runs using a circular window, 45 (22%) detected at least one significant ($P < 0·05$) cluster. An increased sensitivity was found using cases aggregated into 7-day periods: a total of 101 significant clusters were found compared to 70 using a 1-day period. Only two specific clusters were detected by 1-day aggregation models that were not also found using a 7-day aggregation model for the same run date. However, both of these clusters were also contained within larger clusters which were detected as being significant in the 7-day aggregation model. In order to determine the number of clusters likely to have required epidemiological investigation, further analysis was restricted to results from the 7-day aggregation for clusters with at least one circle with a P value $< 0·05$ which remained significant at this level following the removal of household cases. This resulted in 20 clusters for analysis and corresponds to an average of one cluster for epidemiological investigation every 10 weeks (Table 1). The detection of clusters does not appear to be associated with seasonal peaks in incidence (Fig. 1).

All but two of the clusters occurred within the centre-east area of the North East region of England which corresponds with the highest population density (Fig. 3). The highest probability circle (HPC) for each of these clusters was used for comparison of cases (Table 2). The median number of cases within the HPC for each cluster was 30 [interquartile range (IQR) 15–66] with a median population coverage of 5·3% (IQR 2·3–17·6) or a population of 134 442 (IQR 58 250–449997). In total, 928 cases were found within the 20 clusters, representing 7·5% of the total cases reported for 2008–2011.

Similar results were obtained using an elliptical cluster detection window. There was agreement (in terms of the significant detection of the same cluster by the circular and elliptical window models) for 196/209 (93·8%) weeks with discordance due to a lack of statistical significance ($0·054 < P < 0·342$) for either the circular window model (eight runs) or the elliptical window model (five runs) (results not shown). This was not associated with a significantly increased sensitivity for the elliptical window model (48/209 weeks compared to 45/209 weeks for the circular window model; binomial probability test, $P = 0·738$).
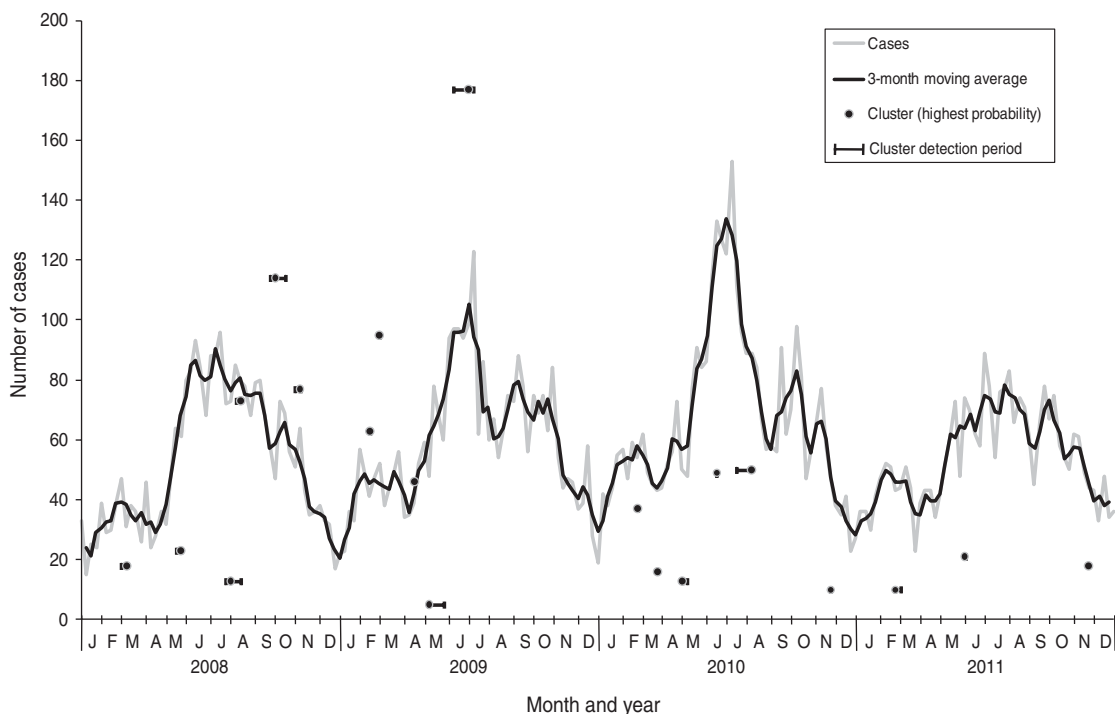
Fig. 1. Laboratory-confirmed *Campylobacter* spp. cases and space-time clusters detected through simulation of weekly prospective detection in the North East of England, 2008–2011. Clusters are positioned horizontally at the week where the cluster was detected with the highest probability and positioned vertically according to the number of cases within the cluster at that time. Horizontal bars indicate the time period where clusters were detected as significant ($P < 0.05$).
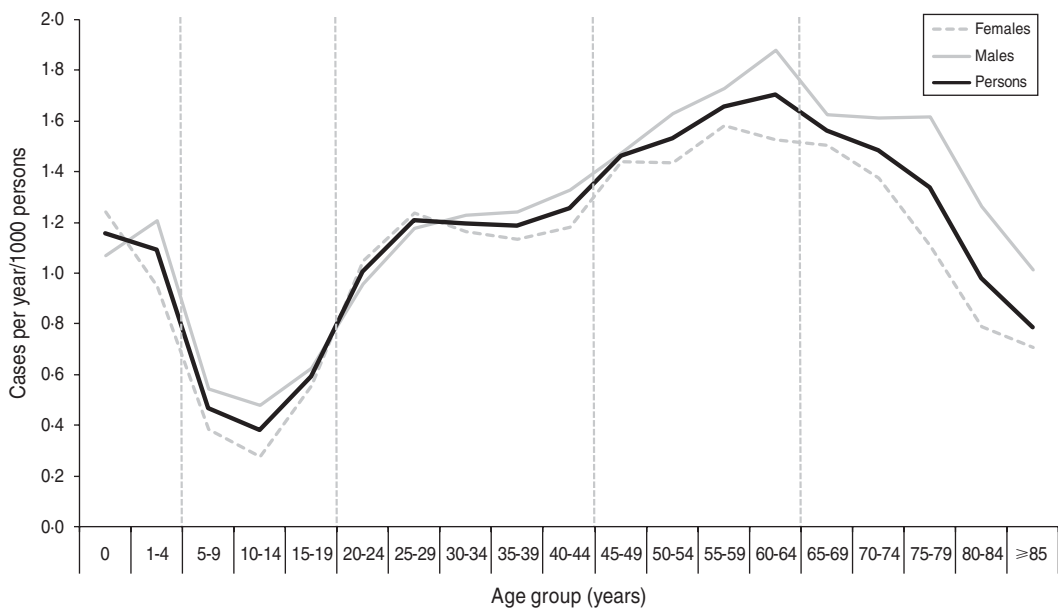

Fig. 2. Incidence rate of laboratory-confirmed *Campylobacter* spp. cases by quinary age groups and sex in the North East of England, 2008–2011. Vertical broken lines represent divisions between age groups used for this study. The figure contains an additional 107 cases from 2007 which were included in the study dataset for simulated runs for January 2008.

Table 1. *Campylobacteriosis clusters detected through simulation of prospective detection in the North East of England, 2008–2011*

Cluster code	First detection (run date)	Start date	Initial largest circle (all cases)†			Largest circle (all cases)‡			Highest probability circle§				
			Cases (expected)	Population (%)*	P value	Cases (expected)	Population (%)*	P value	Duration (weeks)	Cases (expected)	Relative risk	Population (%)*	P value
2008/1	25 Feb. 2008	11 Feb. 2008	3 (<1)	0·1	0·042	16 (4)	6·8	0·036	4	18 (5)	4·02	4·4	0·034
2008/2	12 May 2008	14 Apr. 2008	30 (12)	12·1	0·026	As initial largest circle			4	23 (6)	4·08	5·3	0·003
2008/3	21 July 2008	7 July 2008	9 (1)	0·9	0·041	13 (2)	0·9	0·001	3	13 (2)	7·51	0·9	0·001
2008/4	28 July 2008	30 June 2008	23 (7)	2·8	0·035	121 (83)	34·1	0·038	4	73 (38)	2·23	15·5	0·001
2008/5	22 Sept. 2008	25 Aug. 2008	167 (122)	48·9	0·006	As initial largest circle			4	114 (69)	2·05	28·7	<0·001
2008/6	27 Oct. 2008	6 Oct. 2008	63 (35)	24·8	0·044	As initial largest circle			4	77 (46)	1·99	23·8	0·034
2009/1	9 Feb. 2009	12 Jan. 2009	70 (43)	37·8	0·030	As initial largest circle			4	63 (36)	2·32	32·4	0·026
2009/2	23 Feb. 2009	26 Jan. 2009	100 (69)	49·1	0·024	As initial largest circle			4	95 (60)	2·27	43·4	0·002
2009/3	13 Apr. 2009	16 Mar. 2009	As highest probability circle			As highest probability circle			4	46 (22)	2·42	15·4	0·022
2009/4	4 May 2009	27 Apr. 2009	5 (<1)	0·7	0·032	116 (82)	49·9	0·042	1	5 (<1)	71·83	0·2	<0·001
2009/5	8 June 2009	11 May 2009	137 (97)	48·8	0·006	As initial largest circle			3	177 (118)	2·07	44·3	<0·001
2010/1	15 Feb. 2010	18 Jan. 2010	As highest probability circle			As highest probability circle			4	37 (16)	2·66	11·7	0·025
2010/2	15 Mar. 2010	15 Feb. 2010	As highest probability circle			As highest probability circle			4	16 (4)	4·49	2·3	0·049
2010/3	19 Apr. 2010	22 Mar. 2010	As highest probability circle			29 (12)	9·5	0·047	4	13 (3)	5·32	2·2	0·033
2010/4	7 June 2010	10 May 2010	92 (59)	28·6	0·040	As initial largest circle			3	49 (23)	2·44	11·1	0·006
2010/5	5 July 2010	7 June 2010	132 (90)	25·7	0·044	As initial largest circle			4	50 (20)	2·61	5·2	0·001
2010/6	15 Nov. 2010	18 Oct. 2010	As highest probability circle			As highest probability circle			3	10 (1)	7·69	0·7	0·032
2011/1	14 Feb. 2011	24 Jan. 2011	15 (3)	3·7	0·013	19 (5)	3·8	0·024	3	10 (1)	9·61	1·3	0·004
2011/2	23 May 2011	2 May 2011	As highest probability circle			As highest probability circle			3	21 (5)	4·41	5·1	0·002
2011/3	14 Nov. 2011	17 Oct. 2011	As highest probability circle			As highest probability circle			4	18 (5)	4·00	2·8	0·048

* Percentage of the total population of the study region.

† If different from the highest probability circle.

‡ If different from the initial largest circle and the highest probability circle.
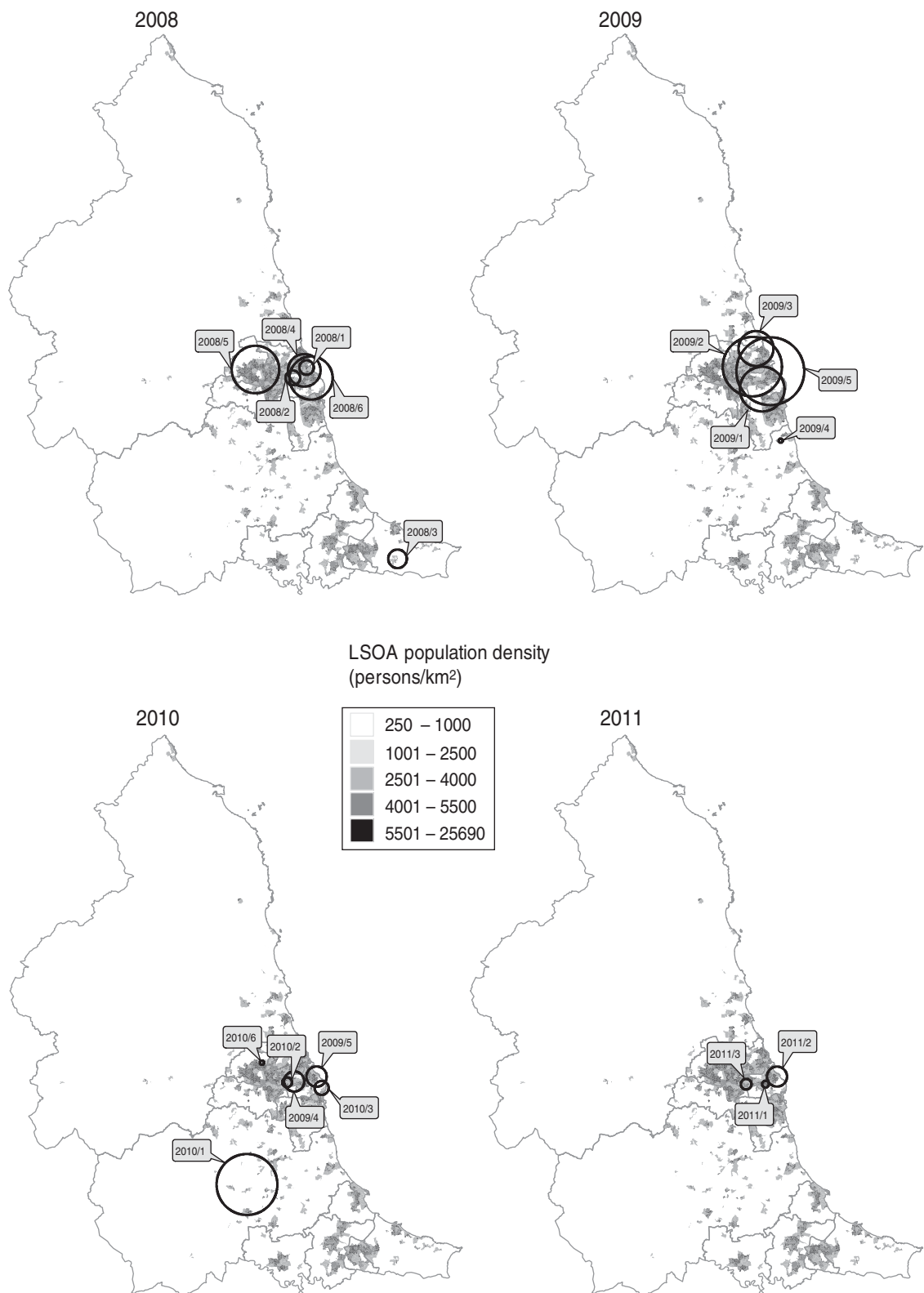
§ After removal of household cases.

Fig. 3. Location of highest probability circles for space-time clusters of laboratory-confirmed *Campylobacter* spp. cases in the North East of England detected through simulated prospective analysis, 2008–2011. Labels indicate the cluster code assigned in this study. Boundaries indicate current primary-care organizations. LSOA, Lower super output area.

Table 2. *Epidemiological details of the highest probability circle for clusters of campylobacteriosis detected through simulated prospective detection in the North East of England, 2008–2011*

Cluster code	Age Median (range) Clustered	Elsewhere	Age Fisher's exact P value	Age group distribution Different groups*	Sex Percentage female Clustered	Elsewhere	Sex Fisher's exact P value	Indices of multiple deprivation 2010 Median (interquartile range) Clustered	Elsewhere	Exact permutation test (P value)	Urbanity classification† 5	6	7
2008/1	56 (3–81)	49 (1–87)	0·4799	≥65	55·6	34·9	0·1164	25·1 (12·3–38·4)	24·1 (12·8–38·9)	0·908	76	—	—
2008/2	51 (17–75)	39 (0–90)	0·1547	≥65	30·4	55·4	0·0381	36·5 (28·1–43·0)	22·9 (12·4–38·3)	0·002	92	—	—
2008/3	50 (2–76)	46 (0–90)	0·4566	—	46·2	48·5	1·0000	11·3 (5·9–22·0)	24·5 (12·9–39·0)	0·012	12	3	—
2008/4	44 (0–90)	45 (1–90)	0·1881	—	41·1	45·7	0·5671	26·4 (12·5–39·5)	23·7 (12·9–38·8)	1·000	264	3	—
2008/5	46 (0–87)	38 (0–81)	0·0727	—	52·6	45·1	0·2145	25·1 (12·7–38·6)	23·7 (12·9–39·1)	0·610	434	45	1
2008/6	48 (0–84)	44 (0–88)	0·3866	—	42·9	47·9	0·5566	28·1 (14·6–40·9)	22·5 (12·1–38·1)	0·012	406	3	—
2009/1	36 (0–81)	47 (1–80)	0·0855	—	57·1	36·8	0·0224	27·1 (13·5–40·0)	22·4 (12·1–38·1)	0·050	542	4	—
2009/2	44 (0–82)	46 (1–88)	0·6761	—	56·8	47·9	0·5296	25·7 (13·2–39·8)	22·4 (12·4–38·1)	0·296	700	29	—
2009/3	54 (5–85)	43 (0–89)	0·2428	—	60·9	46·2	0·2202	22·4 (11·2–34·6)	24·8 (13·1–39·6)	0·002	240	18	—
2009/4	59 (55–75)	43 (2–68)	0·2679	45–64	40·0	58·3	0·6199	36·8 (33·4–42·5)	24·1 (12·8–38·9)	0·288	—	3	—
2009/5	43 (0–88)	48 (1–94)	0·1031	—	38·4	45·5	0·3073	26·2 (13·3–39·6)	22·3 (12·0–38·2)	0·178	712	29	—
2010/1	51 (0–88)	45 (0–84)	0·7527	—	59·5	47·4	0·4582	24·8 (12·9–37·3)	24·0 (12·8–39·4)	0·192	105	71	12
2010/2	50 (17–85)	50 (0–90)	0·3356	5–19	62·5	51·0	0·4382	32·3 (20·5–41·7)	23·8 (12·7–38·9)	0·038	39	—	—
2010/3	62 (0–72)	47 (0–83)	0·0836	≥65	69·2	55·8	0·5512	18·0 (10·2–31·6)	24·5 (12·9–39·1)	0·022	38	—	—
2010/4	54 (1–83)	48 (0–89)	0·6598	—	51·0	48·8	0·5094	33·2 (20·3–42·5)	22·6 (12·3–38·1)	0·002	191	—	—
2010/5	46 (1–83)	48 (0–92)	0·8612	—	52·0	50·5	0·8799	25·1 (12·1–40·7)	24·0 (12·8–38·8)	0·952	90	—	—
2010/6	46 (1–78)	47 (0–94)	0·1524	—	40·0	50·6	0·7463	17·4 (5·4–19·6)	24·5 (12·8–39·0)	0·042	10	—	—
2011/1	51 (12–76)	50 (0–75)	0·6547	—	40·0	45·6	1·0000	37·4 (28·6–42·4)	23·8 (12·8–38·7)	0·038	23	—	—
2011/2	48 (0–72)	43 (1–84)	0·5381	—	57·1	54·2	1·0000	25·3 (13·0–40·5)	24·0 (12·8–38·8)	0·862	87	—	—
2011/3	47 (0–83)	49 (0–86)	0·2824	—	50·0	49·3	1·0000	32·3 (20·6–40·3)	23·7 (12·7–38·9)	0·034	47	—	—

* Age groups with significantly different numbers of observed cases compared to simulated age distributions.
† Number of lower super output areas within the highest probability circle; zero values have been omitted for ease of presentation. 5, Urban settlement with >10 000 population; 6, small towns and fringe areas that are located within the rural domain; 7, villages that are located within the rural domain.

Comparison of clustered and non-clustered cases

Although no age distribution (by age group) for clustered cases compared to the non-clustered cases differed significantly according to the Fisher's exact test, simulated age group distributions suggested significant differences for five of the 20 clusters (Table 2). Two clusters had significantly different expected numbers of cases by sex; one of which was associated with a higher number of males and one of which was associated with a higher number of females (Table 2). The IMD for those LSOAs within clusters significantly differed from the distribution of indices for other LSOAs in the region for 10 of the 20 clusters (Table 2). Six of these differences were associated with a median lower deprivation index (less deprived) within the cluster window and four with a median higher deprivation index (more deprived) within the cluster window. The urbanity classification for cluster LSOAs tended to fit to the general pattern for the North East region as a whole (81 % RUAC of 5, 13 % RUAC of 6).

DISCUSSION

The epidemiology of endemic and relatively prevalent pathogens is such that the detection of periods of increased incidence using time-series analysis requires both a rigid, pre-defined geographical unit (usually a testing laboratory or residential area) and historical baseline incidence data in order to place current epidemiological data in a reliable computational context amid possible seasonal variation and/or periodicity [29]. A review of software for space-time cluster detection discussed alternative methods and recommended SaTScan for use with surveillance systems, particularly where automation may be a long-term goal [21]. The great advantage of space-time cluster detection methods such as SaTScan is the lack of reliance upon strictly bounded areas (within which exceedances are monitored) and the removal of the need to adjust for seasonality [24, 30], although if using the Poisson model there still remains the need to define geographical units (with population data) within the study region.

The integration of SaTScan into existing surveillance systems is certainly possible and is potentially automatable. The surveillance data itself requires little modification prior to running models. All of the geographical and population data used here are freely available. Outside of our setting, appropriate population data is often available and centroids can be readily calculated for geographical areas if they are not already available. Although we primarily report cluster detection using a circular window, use of a more flexible but computationally more expensive elliptical window is possible [23]. However, we found very similar results for both window shapes which is not unsurprising given the relatively high level of geographical aggregation of our analysis (e.g. LSOA compared to postcode of residence) and the lack of evidence or intuition to suggest that *Campylobacter* cases associated with a common exposure would be necessarily distributed in a markedly un-circular manner. Moreover, the propensity for elliptical window models to detect more eccentrically shaped [23] and potentially spurious clusters could be unhelpful to public health resources. Whichever model is used it is important to reiterate that the results of cluster detection serve solely as an indication of a situation requiring investigation [24] and not as a means to define epidemiologically linked cases.

Certainly, a degree of caution should be observed when applying these models and interpreting their output. The methodology underlying the models is complex [22] and the probability assigned to a cluster should not be inferred as a measure of the likelihood of epidemiological linkage prior to an assessment of the cases themselves [24]. Potential local heterogeneity of confounding factors such as the incidence of travel-associated infection and the proportion of symptomatic patients who have a specimen taken for laboratory testing, may also create a source of error which must be considered prior to undertaking any public health action. Potential bias may also occur when population data lags behind population change, as could occur when substantial new housing is created within a geographical unit. A further consideration is that this method would not detect a rise in incidence affecting an entire study region evenly; under these circumstances time-series analysis based on historical data is better suited [29]. Within the context of this study we also used home postcode as a proxy for the location of exposure, which will lead to a reduced sensitivity for detecting outbreaks where the likelihood of exposure is not proportional to the proximity of residential address to the infection source. It is also feasible that some systematic (but yet unknown) bias in the interval between the effective exposure to *Campylobacter* and our proxy measure of this event could lead to the inconsistent placement of cases into temporal windows. However, formal

consideration of this bias would be complex and we feel it unlikely to have resulted in a high type I error rate for this study.

Fundamentally, for the current study we asked whether or not space-time cluster detection could provide a platform upon which suggested clusters of campylobacteriosis could be investigated for the presence of epidemiologically linked cases. Ideally, our results would have been validated against a list of known outbreaks. However, as mentioned earlier, recognized outbreaks of *Campylobacter* tend only to be associated with the catering industry [16]. Under those circumstances, laboratory confirmation may be limited to the first few cases and as such the signal for outbreaks of this kind within routine laboratory surveillance data can be dampened. Nonetheless, despite the lack of historical outbreak data to validate the clusters detected here we did find limited evidence from age distributions and characteristics of cluster areas to suggest that clustered cases are not simply representative of background cases. It remains unclear whether or not the clusters detected here represent outbreaks or whether they are in fact artefacts of cluster detection created through a combination of the complex factors mentioned above. However, should the degree of historical clustering found here be applicable to the contemporary epidemiology of campylobacteriosis in the North East of England then neither the number, size, or geographical spread are restrictive to the application of further epidemiological and microbiological investigation. Without such an approach it will always be impossible to disentangle the complex possibilities of travel-associated exposures [6] environmental exposures [8–10, 30] and food-associated point-source outbreaks [4, 16].

The scale and approach of any further investigation of space-time clusters will partially depend on the number of clustered cases. This is predicted here to vary from less than five to greater than 150, although the expected median of 30 cases every 10 weeks should be manageable by a regional epidemiology unit in England. Following the decision to investigate a cluster of *Campylobacter* cases there are a number of possible approaches that could be taken to perform the actual investigation and a number of means by which data could be obtained, many of which would depend on the local working environment. A descriptive study of those clustered cases would perhaps be the most straightforward approach although a case-case (*vs.* un-clustered cases) or case-control study would probably be more informative and allow direct hypothesis testing. Whatever approach is taken, the findings of this study support the feasibility of a prospective trial of campylobacteriosis cluster detection with further investigation of detected clusters for common links or risk factors.

ACKNOWLEDGEMENTS

DECLARATION OF INTEREST

None.

REFERENCES

1. **Health Protection Agency.** Campylobacter infections per year in England and Wales, 2000–2010. 2011 (http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/Campylobacter/EpidemiologicalData/campyDataEw/). Accessed 14 May 2012.

2. **Nylen G, *et al.*** The seasonal distribution of *Campylobacter* infection in nine European countries and New Zealand. *Epidemiology and Infection* 2002; **128**: 383–390.

3. **Kovats RS, *et al.*** Climate variability and campylobacter infection: an international study. *International Journal of Biometeorology* 2005; **49**: 207–214.

4. **Rodrigues LC, *et al.*** The study of intestinal disease in England: risk factors for cases of infectious intestinal disease with *Campylobacter jejuni* infection. *Epidemiology and Infection* 2000; **127**: 185–193.

5. **Domingues AR, *et al.*** Source attribution of human campylobacteriosis using a meta-analysis of case-control studies of sporadic infections. *Epidemiology and Infection* 2012; **140**: 970–981.

6. **Zenner D, Gillespie I.** Travel-associated *Salmonella* and *Campylobacter* gastroenteritis in England: estimation of under-ascertainment through national laboratory surveillance. *Journal of Travel Medicine* 2011; **18**: 414–417.

7. **Stanley K, Jones, K.** Cattle and sheep as reservoirs of *Campylobacter*. *Journal of Applied Microbiology* 2003; **94**: 104S–113S.

8. **Jonsson ME, *et al.*** Analysis of simultaneous space-time clusters of *Campylobacter* spp. in humans and in broiler flocks using a multiple dataset approach. *International Journal of Health Geographics* 2010; **9**: 48.

9. **Nygard K, *et al.*** Association between environmental risk factors and *Campylobacter* infections in Sweden. *Epidemiology and Infection* 2004; **132**: 317–325.

10. **Potter RC, Kaneene JB, Hall WN.** Risk factors for sporadic *Campylobacter jejuni* infections in rural Michigan: a prospective case-control study. *American Journal of Public Health* 2003; **93**: 2118–2123.

11. **Sinton LW, *et al.*** Survival of indicator and pathogenic bacteria in bovine feces on pasture. *Applied and Environmental Microbiology* 2007; **73**: 7917–7925.

12. **Cook KL, Bolster CH.** Survival of *Campylobacter jejuni* and *Escherichia coli* in groundwater during prolonged starvation at low temperatures. *Journal of Applied Microbiology* 2007; **103**: 573–583.

13. **Hansson I, *et al.*** Correlations between *Campylobacter* spp. prevalence in the environment and broiler flocks. *Journal of Applied Microbiology* 2007; **103**: 640–649.

14. **Thomas C, Hill DJ, Mabey, M.** Evaluation of the effect of temperature and nutrients on the survival of *Campylobacter* spp. in water microcosms. *Journal of Applied Microbiology* 1999; **86**: 1024–1032.

15. **Wallace JS, *et al.*** Seasonality of thermophilic *Campylobacter* populations in chickens. *Journal of Applied Microbiology* 1997; **83**: 219–224.

16. **Little CL, *et al.*** A recipe for disaster: outbreak of campylobacteriosis associated with poultry live pâté in England and Wales. *Epidemiology and Infection* 2010; **138**: 1691–1694.

17. **Inns T, *et al.*** Cohort study of a campylobacteriosis outbreak associated with chicken liver parfait, United Kingdom, June 2010. *Eurosurveillance* 2010; **15**: pii = 19704.

18. **Rooney R, *et al.*** Survey of local authority approaches to investigating sporadic cases of suspected food poisoning. *Communicable Disease and Public Health* 2000; **3**: 101–105.

19. **Gillespie IA, *et al.*** Point source outbreaks of *Campylobacter jejuni* infection – are they more common than we think and what might cause them? *Epidemiology and Infection* 2003; **130**: 367–375.

20. **Kulldorff M.** Information Management Services, Inc. SaTScan™ v8.0: software for the spatial and space-time scan statistics (www.satscan.org).

21. **Robertson C, Nelson TA.** Review of software for space-time disease surveillance. *International Journal of Health Geographics* 2010; **9**: 16.

22. **Kulldorff M.** A spatial scan statistic. *Communications in Statistics: Theory and Methods* 1997; **26**: 1481–1496.

23. **Kulldorff M, *et al.*** An elliptic spatial scan statistic. *Statistics in Medicine* 2006; **25**: 3929–3943.

24. **Kulldorff M.** Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 2001; **164**: 61–72.

25. **Jones RC, *et al.*** Use of a prospective space-time scan statistic to prioritize shigellosis case investigations in an urban jurisdiction. *Public Health Reports* 2006; **121**: 133–139.

26. **Department for Communities and Local Government.** English indices of deprivation 2010 (http://www.communities.gov.uk/documents/statistics/pdf/1871538.pdf). Accessed 14 May 2012.

27. **Office for National Statistics.** Rural and urban Statistics in England: guidance notes (http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/rural-urban-definition-and-la/rural-urban-definition-england-and-wales-/rural-and-urban-statistics-guidance-notes.pdf). Accessed 14 May 2012.

28. **R Development Core Team.** R: a language and environment for statistical computing. R Foundation for Statistical Computing (http://www.R-project.org). Accessed 16 May 2012.

29. **Farrington CP, *et al.*** A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 1996; **159**: 547–563.

30. **Duke LA, *et al.*** A mixed outbreak of *Cryptosporidium* and *Campylobacter* infection associated with a private water supply. *Epidemiology and Infection* 1996; **116**: 303–308.