
Informing surveillance programmes by investigating spatial dependency of subclinical *Salmonella* infection

J. BENSCHOP^{1*}, M. A. STEVENSON¹, J. DAHL², R. S. MORRIS¹
AND N. P. FRENCH¹

¹ EpiCentre, Institute of Veterinary, Animal, and Biomedical Sciences, Massey University, Palmerston North, New Zealand

² Danish Meat Association, Axeltorv, Copenhagen, Denmark

(Accepted 11 January 2009; first published online 19 February 2009)

SUMMARY

The aim of this paper is to investigate local spatial dependency with regard to *Salmonella* seropositivity in data from the Danish swine salmonellosis control programme and its application in informing surveillance strategies. We applied inhomogeneous and observed-difference *K*-function estimation, and geo-statistical modelling to data from the Danish swine salmonellosis control programme. Slaughter-pig farm density showed large variation at both the country-wide and local level in Denmark (median 0.23, range 0.02–0.47 farms/km²). The spatial distribution of pig farms followed a random inhomogeneous Poisson process but was not aggregated. We found evidence for aggregation of *Salmonella* case farms over that of all farms at distances of up to 6 km and semivariogram analyses of *Salmonella* seropositivity revealed spatial dependency between pairs of farms up to 4 km apart. The strength of the spatial dependency was positively associated with slaughter-pig farm density. We proposed sampling more intensively those farms within a 4 km radius of farms that were identified with a high *Salmonella* status, and reduced sampling of farms that are within this radius of ‘*Salmonella*-free’ farms. Our approach has the potential to optimize sampling strategies while maintaining consumer confidence in food safety and also has potential to be used for other zoonotic disease surveillance systems.

Key words: *K* function, *Salmonella*, spatial, surveillance, variogram.

INTRODUCTION

The value of geo-referenced data in veterinary surveillance of both endemic and exotic diseases is immense. Recent examples in the literature show that these data have been used not only to identify areas with excess disease [1, 2] and target areas for further studies [3], but also to produce hypotheses about means of disease introduction [4], identify likely sites

of incursion of an exotic disease [5] and for predictive modelling of alternative control strategies [6].

The geo-referenced locations of livestock farms can be considered a spatial point process [7]. An underlying assumption in the analysis of these processes is that of stationarity or spatial homogeneity, i.e. the intensity of the process does not depend on the location in space [7, 8]. A point pattern representing the location of livestock farms will typically not meet this assumption – farms will probably be distant from large urban centres and will often be located near areas that meet their needs for specific inputs, e.g. feed supply and market access. Furthermore, in developed

* Author for correspondence: Ms. J. Benschop, EpiCentre, Institute of Veterinary, Animal, and Biomedical Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand. (Email: j.benschop@massey.ac.nz)

countries legislation now dictates the location of intensive production units due to their effects on the environment such as emissions of ammonia and phosphorus and requirements regarding the spread of slurry.

Statistically spatial point patterns can be partitioned into first- and second-order properties that capture their global and local behaviours respectively [8]. If the pattern shows a global trend (i.e. is non-stationary or inhomogeneous) then it exhibits a first-order effect. A second-order effect is due to spatial dependency and results from the spatial correlation structure in the data; these are small-scale or local effects. Somewhat ambiguously, both first- and second-order effects produce point patterns that exhibit local concentrations of points and it can be difficult to clearly identify one from the other [9].

Specifically viewing slaughter-pig production in intensive farming areas there are concentrated areas of pig production within which the distances between farms can be very small. Denmark, as the world's largest exporter of pig meat provides a good example of intensive pig farming. The first aim of this paper is to capture the spatial distribution of these farms with regard to first- and second-order effects using farm location data from the Danish Central Husbandry Register in 2003. Our second aim is to investigate the second-order spatial properties by marking the locations with disease status and with a random farm-effect value from a generalized linear mixed model. We then determined the implications for surveillance. This methodology could be used on suitable data from any national disease control programme. We used data from the Danish Swine Salmonellosis Control Programme (DSSCP) from 2003. Many other countries that intensively farm pigs look to the Danish control programme as a model, e.g. the Zoonoses Action Plan in the UK [10], Ireland [11] and the German QS system [12]. The Danish programme was developed in 1993 in response to an increase in the incidence of salmonellosis in humans attributable to consumption of pork [13, 14] and is based around the random testing of meat-juice samples from slaughtered pigs. All herds that produce >200 finishers per year are tested and then categorized into one of three levels of a 'serological *Salmonella* index' for intervention strategies [13]. An in-depth review of the programme is given by Christensen [15].

In Denmark the number of human salmonellosis cases due to pork consumption has substantially reduced from 1444 in 1993 to 164 in 2004 [16]. This

reduction in the number of human cases provides some indication that interventions that have been applied have been effective but raises questions about where to go to next in terms of resource allocation within the programme [17]. There have been a number of recent stochastic models, both Danish [17] and from elsewhere [18, 19] which have addressed the question, with variable results. The North American model found higher cost-benefit ratios for improvements in the post-slaughter phase [19], while both the Danish [17] and the Dutch [18] models identified both pre- and post-slaughter interventions as being efficient.

In terms of pre-slaughter interventions little consideration has been given to small-scale spatial risk factors. Work on the Danish programme has described a strong first-order spatial effect with a higher prevalence of farm-level seropositivity in the north and south of Jutland, and in the west of the country compared with the east [14, 20, 21]. Our recent work [21] has identified that case farms tend to be spatially aggregated, but we are not aware of any work specifically investigating the second-order properties of the data. Both increased pig density within a region [22] and small distances to other pig farms [23, 24] have been identified as risk factors for *Salmonella* infections. Survival times of bacteria are lengthy in the environment [25] and contaminated faecal matter can act as a reservoir [26], so processes acting locally, such as sharing contaminated agricultural machinery or poor biosecurity between farms, make the small-scale spatial structure worthwhile investigating. This has the potential to inform models that may lead to improved resource allocation in the Danish and other similar programmes.

MATERIALS AND METHODS

The dataset

Two extracts of data from 1 January 2003 to 31 December 2003 were obtained from the DSSCP [13, 14]. These extracts comprised pig- and farm-level data. We chose data from 2003 for analysis since this was the period with the highest proportion of geo-referenced farms in our dataset (96.2%).

Data were managed using a relational database (Microsoft Access 2002 for Windows; Microsoft Corporation, USA) and spreadsheet software (Microsoft Excel 2002 for Windows; Microsoft Corporation). Statistical analyses were performed using the R statistical package version 2.2.0 (R Foundation;

<http://www.r-project.org>) and WinBUGS version 1.4.1 (Imperial College and MRC, UK). R packages spatstat [27], geoR [28], splancs [29] and sm [30] were used.

Pig-level data

There were 578 268 individual finisher-pig meat-juice results. Each included the date of sampling, the central husbandry register number identifying the farm of origin, and the result of the Danish-mix ELISA. A result of >20 OD% was classified as positive. This is the cut-off for positivity that has been used by the DSSCP since 1 August 2001 [13].

Farm-level data

Of the 10 571 farms for which individual pig results were available, 10 166 had easting and northing coordinates of the farm house. This represented 96.2% of the contributing farms. The 405 farms without coordinate information were excluded from the analyses. Each farm had its central husbandry register number which included a number indicating within which of the 15 Danish counties the farm was located. Because they contributed very few farms, the two counties that constituted the county of Copenhagen were merged.

Spatial analyses

To investigate the spatial distribution of slaughter-pig farms we used three techniques: kernel estimation, nearest-neighbour distance and the inhomogeneous K function.

We calculated kernel density estimates [31] of farm locations to visualize the broad scale variability in farm density. Spatially adaptive smoothing was implemented by weighting the global bandwidth at each data-point with weights derived from a pilot estimate (J. C. Marshall and M. L. Hazelton, unpublished observations). Regions that are data rich (e.g. Jutland), therefore receive less smoothing so as to preserve fine detail, whereas regions where the data are sparse (e.g. Zealand) receive more smoothing. A linear boundary kernel, with a Gaussian base was used to reduce boundary bias, and a global smoothing bandwidth of 17 km was chosen using the normal optimal method [30].

For each county we calculated the distance from every farm location to its nearest neighbour.

We estimated a non-stationary analogue of the standard K function, the inhomogeneous K function [32] to investigate for evidence of local aggregations of pig farms after allowing for their non-uniform density. The K function is defined as the expected number of further points within a distance r of an arbitrary point, divided by the overall density of the points [33].

$$K(r) = \frac{N(r)}{\lambda}, \quad (1)$$

In equation (1) $K(r)$ is the standard K function, $N(r)$ is the expected number of neighbouring farms within a distance r of an arbitrary farm and λ is the farm density. Inhomogeneous K -function analysis was performed using five large, approximately square, areas that included 82% of the sampled farms (Fig. 1). Square areas were chosen to avoid the instability that may be associated with unusual window geometry [34]. Analysis of the whole of the country was prevented by computational and geographical constraints. To reduce the instability due to edge effects Ripley's isotropic corrections were implemented [34]. One hundred simulated realizations of an inhomogeneous Poisson process were generated and the inhomogeneous K functions of these were calculated to produce an envelope around the observed data. This provided a way of testing if the observed pattern of farms is aggregated even after allowing for its non-uniform density. The practical value of the inhomogeneous K function over the standard K function is that the former permits a more global measure of aggregation as it allows for spatial inhomogeneity of the pattern (a varying λ).

To investigate if there were spatial aggregations of case farms over that of all farms the observed-difference K function was calculated. A farm was defined as a case if it had a proportion of positive pigs ≥ 0.4 . We chose this cut-off as it is the cut-off between levels 1 and 2 of the serological *Salmonella* finisher index. If herds are in levels 2 or 3 there are requirements placed upon them, e.g. pen faecal samples must be collected from the herd and there are penalty 'Salmonella deductions' reducing payments to these producers. Approximately 3% of herds were in levels 2 or 3 during 2003.

For each county separate K functions at distances r were calculated for both case farms, $K_{\text{case}}(r)$, and for all farms, $K_{\text{pop}}(r)$, and the observed difference function $D(r)$ was calculated as follows:

$$D(r) = K_{\text{case}}(r) - K_{\text{pop}}(r), \quad (2)$$

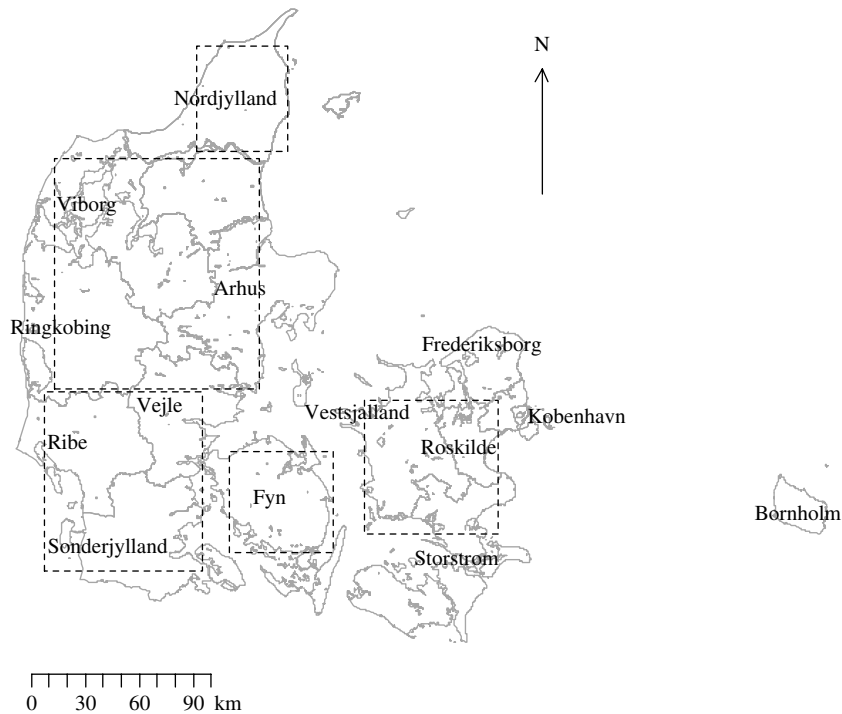


Fig. 1. Map of Denmark showing the location of counties and of the five areas used in the investigation of inhomogeneous K -function estimation.

The null hypothesis was of no extra aggregation of cases over that of the population corresponding to the cases being a random sample from the population. This permits the use of randomization tests which do not require the underlying point process to be stationary [9]. Upper and lower permutation envelopes were produced by 99 random relabellings of the cases and population. Values of the observed-difference function were calculated for each permutation to investigate if there was any significant deviation of the observed-difference function from zero [35].

Our second approach to determine if there were any second-order effects was to investigate the hypothesis that geographically close farms were more similar than those geographically distant. The relationship between the outcome response (the proportion of pigs positive per farm) and the effect of herd size and farm was examined by fitting a generalized linear mixed model as follows:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 x_{ij} + U_i \quad (3)$$

$$U_i \sim \text{Normal}(0, \sigma^2)$$

In Equation (3) the logit of the observed probability of the j th pig from the i th farm being seropositive, p_{ij} , was estimated as a function of a binary variable

representing large herd-size category and a random effect term, U_i , which was normally distributed with a mean of zero and variance σ^2 .

The model was applied to all farms in Denmark that had easting and northing coordinates supplied and were producing pigs for slaughter in 2003. The model was sequentially run for all Danish pig-producing counties as computational constraints prevented modelling all farms at once.

Model parameters were estimated using a Bayesian approach, implemented in WinBUGS version 1.4.1. Markov Chain Monte Carlo (MCMC) methods were applied to the observed data to simulate values from the joint conditional distributions of the unknown quantities. We chose non-informed prior and hyper-prior distributions for all model parameters: for the fixed-effects we chose Normal(0, 0.000001) and for σ^2 (the variance of the farm random-effect term), we chose inverse Gamma(0.1, 0.001). Three chains were run and convergence was judged to have occurred on the basis of visual inspection of time-series and Gelman–Rubin plots [36]. The length of the chain was determined by running sufficient iterations to ensure the Monte Carlo standard errors for each parameter were <5% of the posterior standard deviation. A total of 30 000 iterations were run with a ‘burn in’ of 5000 iterations.

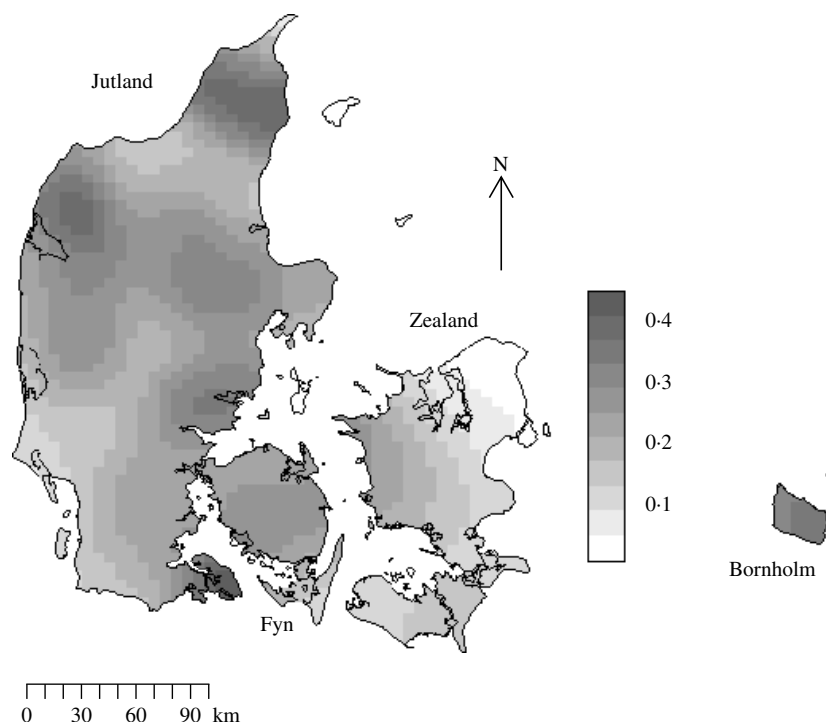


Fig. 2. Kernel-smoothed map showing variation in *Salmonella* meat-juice-tested slaughter herd densities across Denmark. The Jutland peninsula and main islands are labelled. Herds that produced <200 pigs for slaughter annually were not tested. Units are farms/km².

The farm-level random effects from the model were plotted on to county map outlines in an initial investigation into the presence or otherwise of second-order spatial effects. Then omni-directional binned semi-variograms were plotted. These illustrate the difference between pairs of data-points (farm-level random effects) within a given spatial lag (the distance between pairs of farms) [37]. If there was spatial dependency between farms we would expect an upwards trend in the variogram. Conversely, little or no spatial autocorrelation would produce an essentially flat variogram. Directional semivariograms at angle sizes of 0°, 45°, 90° and 135° (tolerance of $\pm 22.5^\circ$) were plotted to investigate if the spatial structure was anisotropic.

The significance of the spatial autocorrelation was determined by permuting the data values on the spatial locations to produce simulation envelopes. As permuted data should not exhibit spatial dependency any points lying outside these simulation envelopes indicate significant spatial autocorrelation. The magnitude of the spatial autocorrelation was determined by calculating the ratio of nugget to total semivariance. The nugget semivariance is the point at which an extrapolated fitted line would cross the vertical axis. A nugget to total semivariance ratio of <25% indicated strong spatial dependence, between 25%

and 75% indicated moderate spatial dependence, and >75% indicated weak spatial dependence [38].

As we were interested in small-scale spatial dependency for both *K* function and semivariogram analysis the maximum distance investigated was 10 km.

RESULTS

There were 10 166 farms sampled in 2003 in the Danish programme with coordinate information. Figure 2 is the edge-corrected kernel-smoothed map of the farm density. Smoothed farm density was normally distributed with a mean of 0.20 and a standard deviation of 0.09 farms/km². The range of smoothed densities varied throughout the country from zero in Copenhagen to 0.47 per in Viborg.

Figure 1 shows the location of counties and the five areas used in the investigation of inhomogeneous *K*-function estimation. Table 1 gives the area, number of farms and farm density for each of the five areas selected for inhomogeneous *K*-function analysis. In total the areas encompassed 8286/10 166 farms sampled for 2003. Over all five areas there was a wide range of farm densities from a median of 0.30 (range 0.03–0.38) farms/km² in North Jutland to 0.14 (range 0.01–0.30) in Zeeland.

Table 1. Area, number of farms, number of case farms and farm density for the five approximately square regions used in *K*-function analysis

Location	Area (km ²)	No. of farms	Farm density (farms/km ²)	
			Median*	Range*
North Jutland	2645	832	0.30	0.03–0.38
Central Jutland	15 960	4232	0.23	0.06–0.39
South Jutland	8712	1623	0.17	0.02–0.32
Fyn	3248	738	0.23	0.05–0.36
Zealand	5625	859	0.14	0.01–0.30

* Calculated using the spatstat library in R.

The inhomogeneous *K*-function analysis of all large square areas showed that the observed pattern of farms was not aggregated (not shown). The median nearest-neighbour distance was 0.77 km (IQR 0.69, range 0.01–11.56 km).

Using the cut-off of $\geq 40\%$ meat-juice, ELISA-positive pigs in a herd produced 272 case farms. The case incidence risk was 3%. Figure 3 shows the observed-difference *K* function between case and population farms for the counties of Nordjylland, Arhus, Ringkøbing and Sønderjylland. Nordjylland, Ringkøbing and Arhus show evidence of local spatial aggregation of case farms over that of all farms. The extent of the aggregation was 1 km for Nordjylland and 4 km for Arhus. For Ringkøbing it was statistically significant at 6 km with points beyond the simulation envelope. Together these three counties represented 40% of the Danish pig population in 2003. The results for the remaining counties were similar to that of Sønderjylland showing no evidence for local spatial aggregation of case farms over that of all farms.

When the farm-level random effects were plotted by their coordinates there were no apparent aggregations of similar-sized random effects. This pattern was seen in all counties. However, semivariograms (Fig. 4) for most large pig-producing counties showed evidence of spatial dependency with an upwards trend in the variogram at up to 4 km distance. Although most counties had all points lying within the simulation envelopes, the four main pig-producing counties Nordjylland, Viborg, Arhus and Sønderjylland had points below the envelopes indicating significant spatial autocorrelation from 2 km to 4 km. Together these four counties represented 50% of the Danish pig population in 2003. The nugget to total semivariance ratios of these four counties was $\sim 70\%$, indicating moderate spatial autocorrelation. The strength of the

dependency was proportional to slaughter-pig density with the exception of Fyn.

Table 2 shows the farm-level prevalence unadjusted for herd size, proportion of farms in the large herd-size category, odds ratios for large herd size and the variance of the random effects with 95% Bayesian credibility intervals for each county. The unadjusted farm-level prevalence was highest at $\sim 5\%$ in the north of Jutland (Nordjylland and Arhus) and lowest, at $\sim 1\%$, in the east of Denmark (Bornholm and Roskilde). All counties in Jutland and Fyn had $\geq 43\%$ farms in the large herd-size category. Odds ratios for Nordjylland, Fyn, Ribe, Vejle and Viborg were significant suggesting that pigs in these counties were at more risk of being seropositive if herd size was large (> 2000 finishers produced annually) than if it was medium (between 200 and 2000 finishers produced annually). The variance of the random effects was greatest in Sønderjylland indicating that farms in this county showed the most variation in farm-level prevalence of *Salmonella*.

DISCUSSION

Slaughter-pig farm density showed large variation both at the country-wide and at the local level in Denmark in 2003. The areas of highest farm density are Viborg and Nordjylland on the Jutland peninsula (0.47 farms/km²); the lowest are on the island of Zealand. The distribution pattern of farms followed a random inhomogeneous Poisson process and although farms had near neighbours they did not spatially aggregate. With regard to *Salmonella* seropositivity we found consistent evidence for spatial dependency at distances of ~ 4 km. The strength of the spatial dependency varied throughout the country being proportional to farm density. Our findings were in accord with those reporting short distances

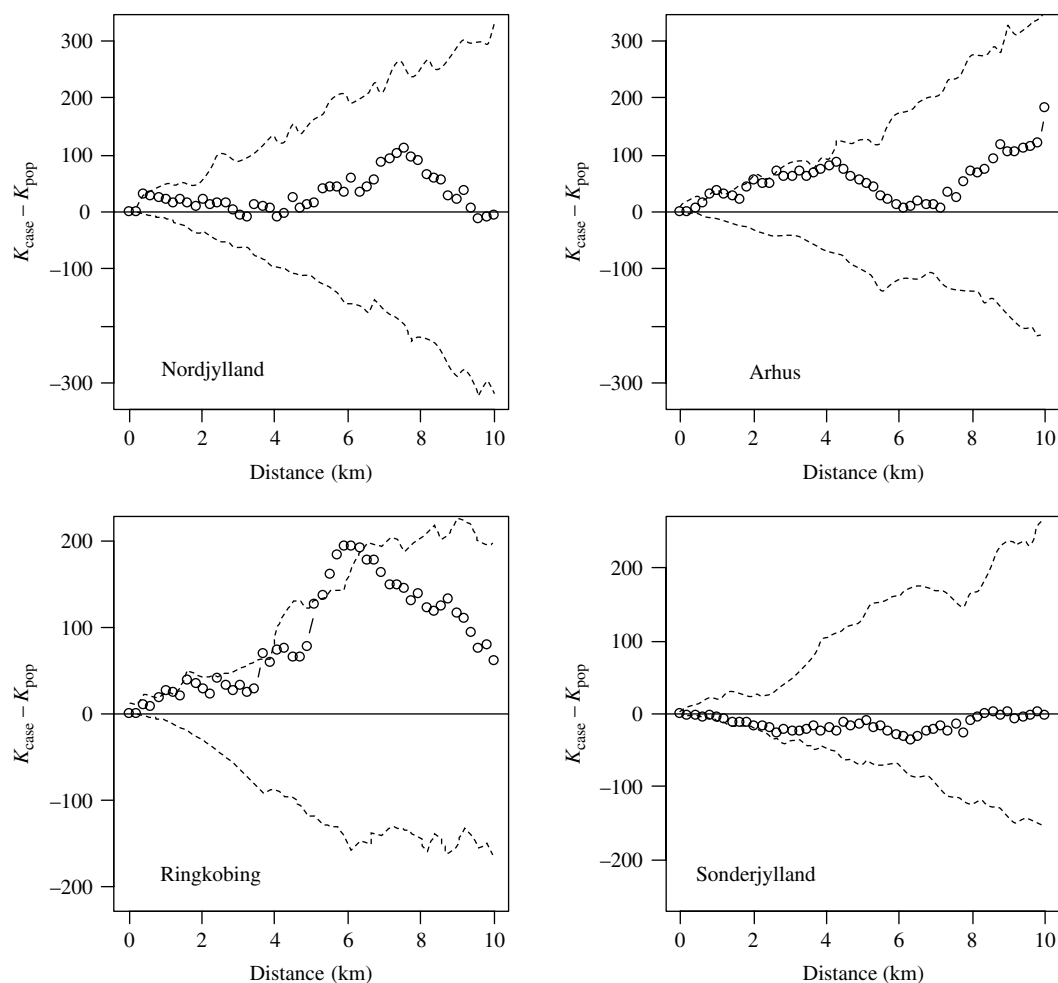


Fig. 3. Observed-difference K function between case and population farms for Nordjylland, Arhus, Ringkobing and Sonderjylland. The open symbols (\circ) represent the difference between the two K functions and the dashed lines (---) the simulation envelope based on 99 random relabellings of the cases and population. A farm was defined as a case if in 2003 the proportion of positive results was $>40\%$.

between farms [23], being a neighbour of an infected farm [24] and pig density [22] as potential risk factors for *Salmonella* infection in pigs.

This local spatial dependency adds to the current knowledge of the epidemiology of subclinical *Salmonella* in Danish slaughter-pig farms and can inform future strategies aimed at optimizing the control programme. For example, more intensive sampling of farms within a 4 km radius of identified problem farms, such as those in levels 2 or 3, on the Jutland peninsula is likely to capture more positive results, leading to interventions that may result in enhanced food safety. Similarly, we propose the concept of reduced sampling of farms that are near neighbours of 'Salmonella-free' farms. 'Salmonella-free' refers to farms enrolled in the 'risk-based' scheme which has been running since July 2005. This scheme requires

one sample per month to be taken from herds with a *Salmonella* index level of 0 and a minimum of 10 negative meat-juice samples in the last 6 months. To date 50% of herds meet these criteria. Our study has identified that when spatial dependency is present, such as in Jutland, there are farms that provide essentially redundant information that could potentially be eliminated from the surveillance programme. Spatial sampling optimization for groundwater monitoring has been achieved using the variogram [39, 40] and we propose using it to optimize sampling in the DSSCP. If spatial dependency is present in other disease programmes, both within and beyond Denmark, then these strategies could be applied to these programmes. For example, evidence for spatial dependency has been found between bulk milk tanks titres for *Salmonella* Typhimurium in Texas dairy herds [3]

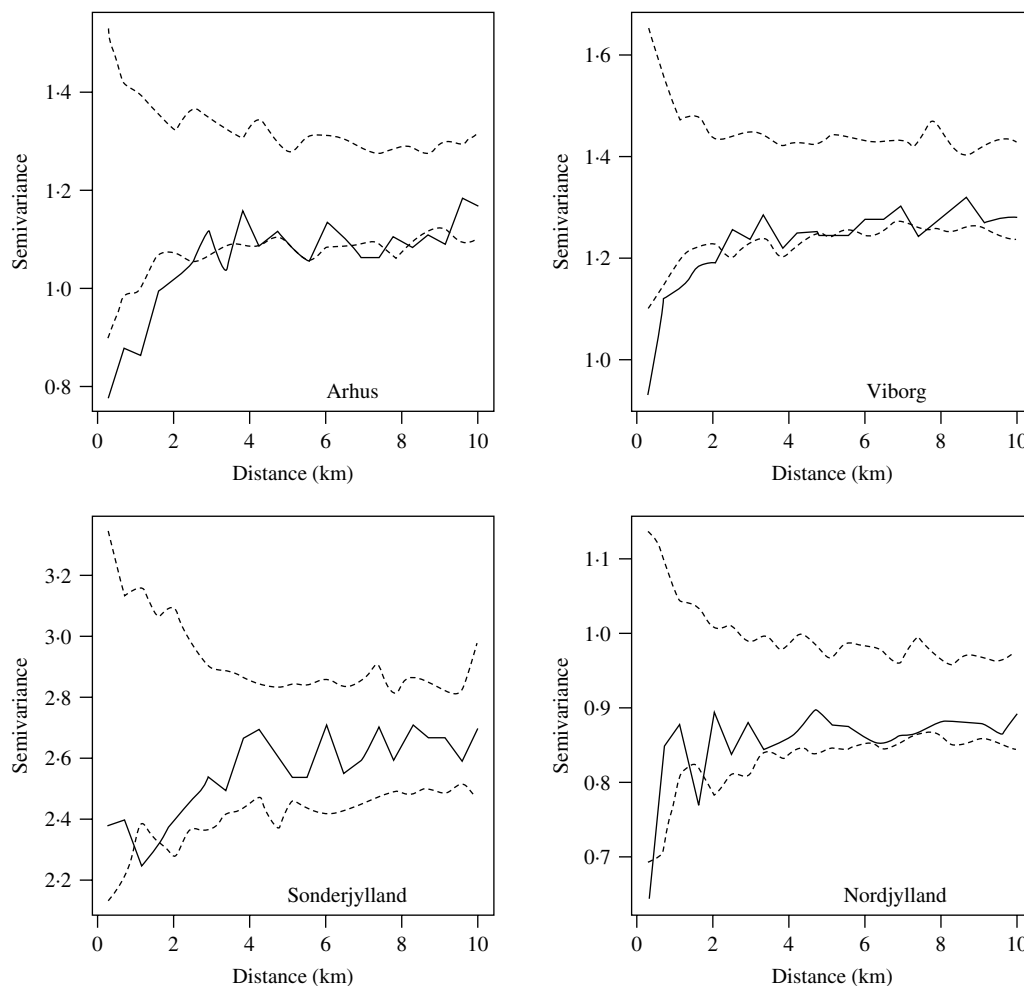


Fig. 4. Spatial semivariograms fitted to the herd-size adjusted farm-level random effects for the counties of Arhus, Viborg, Sonderjylland and Nordjylland. The open symbols (○) represent the semivariance and the dashed lines (---) the simulation envelopes obtained by permutation of the data on the spatial locations.

and between cattle herds in Denmark with regard to *Salmonella* Dublin infection [41].

Local farm density is a well recognized risk factor when investigating epidemics of animal disease [42–45]. The density of neighbouring herds was associated with so-called ‘neighbourhood infections’ during the 1994 classical swine fever epidemic in Belgium [42] and ‘local’ spread accounted for 79% of means of spread in the first 5 months of the 2001 foot-and-mouth disease epidemic in Great Britain [45]. High farm density implies that the distance between farms is short; in these examples a neighbourhood was an area of 1 km radius around an infected herd and local meant within 3 km of an infected place.

Although we are less familiar with farm-density investigation in relation to a subclinical endemic infection such as *Salmonella* in Danish finisher-pig herds there are compelling reasons to investigate it. If

Salmonella is not already present, or if a novel serovar is in circulation, then pig herds are at risk from its introduction through many routes, the two main routes being the introduction of infected pigs and contaminated feed [46]. The latter is thought to be of minor importance as there are stringent controls on animal feed in Denmark; where in 2005 the prevalence of *Salmonella* in animal feed was low. There is much support for the theory that the introduction of infected pigs is a likely source of *Salmonella* for Danish pig farms [23, 47–49]. It is common farming practice to purchase stock from a geographically close supplier and this could lead to small-scale spatial dependency in the data. Denser farming areas probably offer more choice of supplier. Lo Fo Wong *et al.* reported that the odds of seropositivity increased significantly if greater than three suppliers were used [47].

Table 2. Unadjusted farm-level *Salmonella* seroprevalence, odds ratios, proportion of farms in the large herd-size category and variance of the farm-level random effects for Danish pig-producing counties in 2003

County	Prevalence, % (95% CI)	OR (95% CI)	Proportion of farms in large herd-size category*	Variance of the random effects (95% CI)
Nordjylland	5.9 (5.5–6.3)†	1.15 (1.01–1.32)‡	0.40	1.3 (1.2–1.5)
Bornholm	1.0 (0.7–1.3)	0.92 (0.48–1.77)	0.18	2.3 (1.7–3.3)
Sonderjylland	3.3 (2.8–3.8)	1.04 (0.80–1.36)	0.50	3.7 (3.2–4.2)
Fyn	3.6 (3.3–4.0)	1.25 (1.03–1.52)	0.46	1.6 (1.4–1.8)
Viborg	3.5 (3.2–3.9)	1.27 (1.07–1.52)	0.44	2.0 (1.8–2.3)
Storstrom	1.4 (1.2–1.7)	1.33 (0.95–1.85)	0.32	2.0 (1.6–2.5)
Ribe	2.0 (1.6–2.5)	1.66 (1.14–2.41)	0.43	2.8 (2.3–3.5)
Ringkobing	3.5 (3.2–3.9)	0.92 (0.77–1.10)	0.48	2.0 (1.8–2.2)
Arhus	4.9 (4.5–5.4)	1.09 (0.91–1.30)	0.48	1.7 (1.5–1.9)
Roskilde	1.1 (0.7–1.7)	1.65 (0.75–3.75)	0.42	1.7 (1.0–3.2)
Vejle	2.9 (2.5–3.3)	1.34 (1.05–1.70)	0.48	2.2 (1.9–2.5)
Vestsjælland	1.8 (1.5–2.1)	1.01 (0.76–1.34)	0.30	1.8 (1.4–2.1)
Frederiksborg	1.5 (0.8–2.3)	0.87 (0.34–2.26)	0.26	1.6 (0.8–3.2)

OR, Odds ratio; 95% CI, 95% Bayesian credible intervals.

* Farms with >60 pigs tested in 2003 (equates to an annual slaughter of >2000 finishers).

† Unadjusted farm-level prevalence.

‡ Interpretation: In Nordjylland the odds of a pig being seropositive was increased by a factor of 1.15 (95% Bayesian CI 1.01–1.32) if the pig was from a large (>2000 finishers produced annually) herd than if it was from a medium (between 200 and 2000 finishers produced annually) herd.

The other 'external' sources of *Salmonella* such as visitors [50], vermin [22, 51] and sharing of contaminated equipment [24] can also be farm density-dependent. Rodents and flies have been found to carry *Salmonella* [52, 53] and the small distances between many of the Danish pig farms are well within the range of the brown rat [54]. In addition airborne spread is possible at least over short experimental distances [55, 56]. Our findings of spatial dependency between farms with regard to *Salmonella* seropositivity, and aggregation of *Salmonella* case farms over that of all farms at distances of up to 4 km could be due to these locally acting processes or the contagious nature of the disease. Temporal studies would help elucidate this.

The inhomogeneous *K* function is a relatively new technique. It has been used to highlight significant differences in the spatial aggregation of vacuoles in mice brains infected with different transmissible spongiform encephalopathies [57]. The use of the inhomogeneous *K* function to summarize the spatial pattern of farms seems sensible. It allows for the spatial variation in intensity of the underlying point pattern which is likely to occur in animal production systems and is clearly seen in pig farm density in Denmark. By allowing for the non-uniform intensity of the spatial locations of farms it permits hypothesis

testing for aggregation. Our results support the hypothesis that the farm distribution pattern follows a random inhomogeneous Poisson process with no aggregation beyond that.

Even though our dataset was effectively a census of Danish finisher swine herds in 2003 there was potential for selection, misclassification and confounder bias in our study. Selection bias may have occurred when we excluded 405 of 10 571 (4%) of farms because coordinate information was unavailable. As our database was drawn from herds registered in March 2004 the 10 571 farms with available coordinate information were still in production then and were likely to be different from the 405 that no longer were. However, this is likely to be of little importance as this group of farms represents only 4% of the total.

Further selection bias may have occurred in selecting the five large areas for the inhomogeneous *K*-function analysis. These were approximately square and excluded some areas of pig farms (notably Bornholm) and restricted the sites for consideration to those on large land masses. Nonetheless, we believe the coverage of farms within the five areas was suitably representative of all pig farms tested in 2003; 82% were included and the case incidence risk (3%) was the same as that for all farms.

The use of the farmhouse locations over that of the actual polygonal boundaries of the farm may potentially lead to an over-estimation of the distance between farms (misclassification bias). This would be of great significance in extensive sheep or beef cattle farming systems if farm sizes are large. However, it is likely to be of little consequence in intensive production systems such as the Danish pig farms we are investigating in the present study.

The adjustment for herd size in the geostatistical model was made as a number of earlier Danish studies [20, 48] and a recent Canadian study [58] have reported large herd size as a risk factor for increased seropositivity in slaughter-pig herds. However, later studies in Denmark [59] and throughout Europe [47] showed no association and a Dutch study [60] showed that large herd size was protective. In this study the effect of herd size was not investigated *per se*, but adjusting for herd size was undertaken in the context of its effect on spatial dependency. The odds ratios reported suggest increased risk in some counties as herd size increases. The reason for this may be that there are local practices, such as more movements of pigs between farms or higher within-farm pig density, which make large herd size more of a risk for increased *Salmonella* seroprevalence in these counties. Nevertheless these results must be interpreted with caution as the effect of herd size is probably confounded by other covariates, such as feeding and biosecurity practices that we have no information on.

Distance can be defined in different ways; Euclidean, time of travelling or in terms of social networks [61]. Ideally all three definitions should be considered in the spatial epidemiological investigations and we should not constrain 'locality' to only imply spatial proximity. However, this study focused on Euclidean distance between farmhouses but future studies in relation to social networks would appear to be a logical next step. This could be particularly helpful in tracing the dissemination of infected pigs.

We have outlined an approach to combine georeferenced farm location information and routinely collected control programme data using techniques from spatial point pattern and geostatistical analysis. This has extended the current knowledge of the epidemiology of subclinical *Salmonella* in Danish slaughter-pig farms. Furthermore, we have demonstrated how our approach has the potential to optimize sampling strategies while maintaining consumer confidence in food safety. These techniques could be

readily applied to data from other programmes in different countries.

ACKNOWLEDGEMENTS

We thank Adrian Baddeley from the University of Western Australia for his assistance with the inhomogeneous *K*-function analyses. We thank Jonathan Marshall from Massey University for his assistance with the spatially adaptive smoothing.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Haine D, et al.** Herd-level seroprevalence and risk-mapping of bovine hypodermosis in Belgian cattle herds. *Preventive Veterinary Medicine* 2004; **65**: 93–104.
2. **Sanchez J, et al.** Temporal and spatial analysis of the 1999 outbreak of acute clinical infectious bursal disease in broiler flocks in Denmark. *Preventive Veterinary Medicine* 2005; **71**: 209–223.
3. **Graham SL, et al.** Spatial distribution of antibodies to *Salmonella enterica* serovar Typhimurium O antigens in bulk milk from Texas dairy herds. *Preventive Veterinary Medicine* 2005; **69**: 53–61.
4. **Vigre H, et al.** Spatial and temporal patterns of pig herds diagnosed with Postweaning Multisystemic Wasting Syndrome (PMWS) during the first two years of its occurrence in Denmark. *Veterinary Microbiology* 2005; **110**: 17–26.
5. **Stevenson MA, et al.** Spatial epidemiology of the Asian honey bee mite (*Varroa destructor*) in the North Island of New Zealand. *Preventive Veterinary Medicine* 2005; **71**: 241–252.
6. **Yoon H, et al.** Simulation analyses to evaluate alternative control strategies for the 2002 foot-and-mouth disease outbreak in the Republic of Korea. *Preventive Veterinary Medicine* 2006; **74**: 212–225.
7. **Diggle PJ.** *Statistical Analysis of Spatial Point Patterns*, 2nd edn. London: Arnold, 2003, pp. 159.
8. **Banerjee S, Carlin BP, Gelfand AE.** *Hierarchical Modeling and Analysis for Spatial Data*. Florida: Chapman and Hall/CRC, 2004, pp. 452.
9. **Diggle P, et al.** Second-order analysis of inhomogeneous spatial point processes using case-control data. *Biometrics* 2007; **63**: 550–557.
10. **Armstrong D.** Zoonoses action plan *Salmonella* monitoring programme update. *The Pig Journal* 2003; **52** (<http://www.thepigsite.com/pigjournal/volume/52>). Accessed 22 December 2008.
11. **Casey PG, et al.** *Salmonella* carriage in an Irish pig herd: correlation between serological and bacteriological detection methods. *Journal of Food Protection* 2004; **67**: 2797–2800.

12. **Blaha T.** Up-to-date information from the German QS salmonella monitoring and reduction programme. *Deutsche Tierärztliche Wochenschrift* 2004; **111**: 324–326.
13. **Alban L, Stege H, Dahl J.** The new classification system for slaughter-pig herds in the Danish *Salmonella* surveillance-and-control program. *Preventive Veterinary Medicine* 2002; **53**: 133–146.
14. **Mousing J, et al.** Nation-wide *Salmonella enterica* surveillance and control in Danish slaughter swine herds. *Preventive Veterinary Medicine* 1997; **29**: 247–261.
15. **Christensen J.** Danish swine salmonellosis control program: 1993 to 2001. In: Salman MD, ed. *Animal Disease Surveillance and Survey Systems*. Iowa: Iowa State Press, 2003, pp. 185–207.
16. **Nielsen B, et al.** A new *Salmonella* surveillance and control programme in Danish pig herds and slaughterhouses. *Berliner und Munchener Tierärztliche Wochenschrift* 2001; **114**: 323–326.
17. **Alban L, Stärk KD.** Where should the effort be put to reduce the *Salmonella* prevalence in the slaughtered swine carcass effectively? *Preventive Veterinary Medicine* 2005; **68**: 63–79.
18. **van der Gaag MA, et al.** Cost-effectiveness of controlling *Salmonella* in the pork chain. *Food Control* 2004; **15**: 173–180.
19. **Miller GY, et al.** Influence of *Salmonella* in pigs preharvest and during pork processing on human health costs and risks from pork. *Journal of Food Protection* 2005; **68**: 1788–1798.
20. **Carstensen B, Christensen J.** Herd size and seroprevalence of *Salmonella enterica* in Danish swine herds: a random-effects model for register data. *Preventive Veterinary Medicine* 1998; **34**: 191–203.
21. **Benschop J, et al.** Descriptive spatial epidemiology of subclinical *Salmonella* infection in finisher pig herds: application of a novel method of spatially adaptive smoothing. *Veterinary Research* 2008; **39**: 2.
22. **Fedorka-Cray PJ, Gray JT, Wray C.** *Salmonella* infections in pigs. In: Wray C, Wray A, eds. *Salmonella in Domestic Animals*. CABI Publishing, 2000, pp. 191–207.
23. **Berends BR, et al.** Identification and quantification of risk factors in animal management and transport regarding *Salmonella* spp. in pigs. *International Journal of Food Microbiology* 1996; **30**: 37–53.
24. **Langvad B, et al.** Molecular epidemiology of a geographically localized *Salmonella* typhimurium DT104 outbreak in Danish cattle and pigs. *Acta Veterinaria Scandinavica* (Suppl.) 2003; **98**: 237.
25. **Winfield MD, Groisman EA.** Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Applied and Environmental Microbiology* 2003; **69**: 3687–3694.
26. **Gray JT, Fedorka-Cray PJ.** Survival and infectivity of *Salmonella choleraesuis* in swine feces. *Journal of Food Protection* 2001; **64**: 945–949.
27. **Baddeley A, Turner R.** Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 2005; **12**: 1–42.
28. **Ribeiro Jr. PJ, Diggle PJ.** geoR: A package for geostatistical analysis. *R-News* 2001; **1**: 15–18.
29. **Rowlingson B, Diggle PJ.** Splanx: spatial point pattern analysis code in S-Plus. *Computers and Geosciences* 1993; **19**: 627–655.
30. **Bowman AW, Azzalini A.** *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press, 1997, pp. 193.
31. **Diggle PJ.** A kernel method for smoothing point process data. *Applied Statistics* 1985; **34**: 138–147.
32. **Baddeley AJ, Moller J, Waagepetersen R.** Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 2000; **54**: 329–350.
33. **Ripley BD.** The second-order analysis of stationary point processes. *Journal of Applied Probability* 1976; **13**: 255–266.
34. **Ripley BD.** *Statistical inference for Spatial Processes*. Cambridge: Cambridge University Press, 1988, pp. 148.
35. **Chetwynd AG, Diggle PJ.** On estimating the reduced second moment measure of a stationary spatial point process. *Australian and New Zealand Journal of Statistics* 1998; **40**: 11–15.
36. **Toft N, et al.** Assessing the convergence of Markov Chain Monte Carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard. *Preventive Veterinary Medicine* 2007; **79**: 244–256.
37. **Isaaks EH, Srivastava RM.** *An Introduction to Applied Geostatistics*. New York: Oxford University Press, 1989, pp. 561.
38. **Cambardella CA, et al.** Field-scale variability of soil properties in central Iowa soils. *Soil Science Society of America Journal* 1994; **58**: 1501–1511.
39. **Ling M, et al.** Groundwater monitoring plans at small-scale sites – an innovative spatial and temporal methodology. *Journal of Environmental Monitoring* 2003; **5**: 126–134.
40. **Cameron K, Hunter P.** Using spatial models and kriging techniques to optimize long-term ground-water monitoring networks: a case study. *Environmetrics* 2002; **13**: 629–656.
41. **Ersboll AK, Nielsen LR.** The range of influence between cattle herds is of importance for the local spread of *Salmonella* Dublin in Denmark. *Preventive Veterinary Medicine* 2008; **84**: 277–290.
42. **Mintiens K, et al.** Risk analysis of the spread of classical swine fever virus through ‘neighbourhood infections’ for different regions in Belgium. *Preventive Veterinary Medicine* 2003; **60**: 27–36.
43. **Le Menach A, et al.** Modeling spatial and temporal transmission of foot-and-mouth disease in France: identification of high-risk areas. *Veterinary Research* 2005; **36**: 699–712.
44. **Sellers RF.** Comparison of different control strategies for foot-and-mouth disease: a study of the epidemics in Canada in 1951/52, Hampshire in 1967 and Northumberland in 1966. *Veterinary Record* 2006; **158**: 9.
45. **Gibbens JC, et al.** Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Veterinary Record* 2001; **149**: 729–743.

46. **Lo Fo Wong DMA, et al.** Epidemiology and control measures for *Salmonella* in pigs and pork. *Livestock Production Science* 2002; **76**: 215–222.
47. **Lo Fo Wong D, et al.** Herd-level risk factors for sub-clinical *Salmonella* infection in European finishing-pig herds. *Preventive Veterinary Medicine* 2004; **62**: 253–266.
48. **Baggesen DL, et al.** Herd prevalence of *Salmonella enterica* infections in Danish slaughter pigs determined by microbiological testing. *Preventive Veterinary Medicine* 1996; **26**: 201–213.
49. **Stärk KD, et al.** Differences and similarities among experts' opinions on *Salmonella enterica* dynamics in swine pre-harvest. *Preventive Veterinary Medicine* 2002; **53**: 7–20.
50. **Funk JA, Davies PR, Gebreyes W.** Risk factors associated with *Salmonella enterica* prevalence in three-site swine production systems in North Carolina, USA. *Berliner und Munchener Tierarztliche Wochenschrift* 2001; **114**: 335–338.
51. **Steinbach G, Kroell U.** *Salmonella* infections in swine herds-epidemiology and importance for human diseases. *Deutsche Tierarztliche Wochenschrift* 1999; **106**: 282–288.
52. **Letellier A, et al.** Distribution of *Salmonella* in swine herds in Quebec. *Veterinary Microbiology* 1999; **67**: 299–306.
53. **Barber DA, et al.** Distribution of *Salmonella* in swine production ecosystems. *Journal of Food Protection* 2002; **65**: 1861–1868.
54. **Endepols S, et al.** A scheme for the placement of rodenticide baits for rat eradication on confinement livestock farms. *Preventive Veterinary Medicine* 2003; **58**: 115–123.
55. **Oliveira CJ, Carvalho LF, Garcia TB.** Experimental airborne transmission of *Salmonella* Agona and *Salmonella* Typhimurium in weaned pigs. *Epidemiology and Infection* 2006; **134**: 199–209.
56. **Proux K, et al.** Contamination of pigs by nose-to-nose contact or airborne transmission of *Salmonella* Typhimurium. *Veterinary Research* 2001; **32**: 591–600.
57. **Webster S, et al.** Strain-typing transmissible spongiform encephalopathies using replicated spatial data. In: Baddeley A, ed. *Case Studies in Spatial Point Process Modeling*. Berlin: Springer, 2006, pp. 332.
58. **Farzan A, et al.** Prevalence of *Salmonella* spp. on Canadian pig farms using liquid or dry-feeding. *Preventive Veterinary Medicine* 2006; **73**: 241–254.
59. **Stege H, et al.** Data-quality issues and alternative variable-screening methods in a questionnaire-based study on subclinical *Salmonella enterica* infection in Danish pig herds. *Preventive Veterinary Medicine* 2001; **48**: 35–54.
60. **van der Wolf P, et al.** Herd level husbandry factors associated with the serological *Salmonella* prevalence in finishing pig herds in The Netherlands. *Veterinary Microbiology* 2001; **78**: 205–219.
61. **Haining R.** *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press, 2003, pp. 432.