
Temporal and longitudinal analysis of Danish Swine Salmonellosis Control Programme data: implications for surveillance

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, New Zealand*

² *Danish Meat Association, Copenhagen, Denmark*

(Accepted 3 December 2007; first published online 16 January 2008)

SUMMARY

The control programme for *Salmonella* infection in Danish swine has reduced the number of human cases attributable to pork consumption and the focus is now on cost-effectiveness. We applied time-series and longitudinal analyses to data collected between January 1995 and May 2005 to identify if there were predictable periods of risk that could inform sampling strategy; to investigate the potential for forecasting for early aberration detection; and to explore temporal redundancy within the sampling strategy. There was no evidence of seasonality hence no justification to change to targeted sampling at high-risk periods. The forecast of seropositivity made using an ARIMA (0, 1, 2) model had a root-mean-squared percentage error criterion of 8·4% indicating that accurate forecasts are possible. The lorelogram identified temporal redundancy at up to 10 weeks suggesting little value in sampling more frequently than this on the ‘average’ farm. These findings have practical applications for both farm-level sampling strategy and national-level aberration detection which potentially could result in a more cost-effective surveillance strategy.

INTRODUCTION

In industrialized countries most cases of human salmonellosis are foodborne, and pork has been implicated as an important source [1, 2]. In Denmark the estimated number of cases attributed to domestically produced pork has decreased substantially from 22 cases/100 000 head of population in 1993 to 2·6 cases/100 000 in 2004 [3]. This decrease has been largely attributed to the Danish Swine Salmonellosis Control Programme (DSSCP) which was established by the Danish Ministry of Food, Agriculture and Fisheries in 1993.

Finisher pigs are known to carry *Salmonella* and it is these that can contaminate the food product, which is then capable of infecting humans [4]. Carriage occurs in the gut, lymph nodes, and tonsils. Bacteria may directly or indirectly contaminate the carcass during evisceration and other processes that occur in the slaughterhouse such as scalding or cutting [5]. In addition, there is accumulating evidence to suggest that pigs may become infected during transport to slaughter, or while waiting in lairage [6]. However, in Denmark 95% of pigs have less than 3 h of transport and lairage [7, 8], so the chance of infection occurring during these processes is likely to be minimal. Therefore the predominant source of *Salmonella* that contaminates the carcasses and presents a risk to human health will be the farm of origin. What is unknown about the farm-level risk of infection is whether or not

* 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)

it is constant over time or fluctuates on a seasonal basis. Although this issue has been addressed by other authors [9–11] the follow-up period for each of these studies was short.

In this paper our first aim was to apply time-series methods to data collected by the DSSCP over a 10-year period and to describe the key temporal features of trend, cyclicity, and autocorrelation. Subsequent to the successful reduction in pork-attributed human cases attention is now focused on the cost-effectiveness of *Salmonella* surveillance [12] and an understanding of the temporal pattern of *Salmonella* seroprevalence in pigs can partially address this need. If identified predictable periods of high risk are found then a risk-based approach would involve sampling more frequently at these times and less frequently at periods of less risk [13].

Our second aim was to evaluate the predictability of the seroprevalence time series with a view to using the data for forecasting. By comparing forecasts with the data that was actually observed there is the possibility of being able to readily detect aberrant behaviour. The identification of an aberration in national herd seroprevalence could be used as a screening test to alert authorities to emerging problems.

Our third aim is to address the question of temporal redundancy within the sampling strategy. In this study temporal redundancy encompasses the situation where farms are sampled so often that the dependency between these repeated samples could deem some testing unnecessary. We use the lorelogram, a technique related to that previously used in animal tracking [14], environmental radiometry [15], and ground-water monitoring [16] to suggest an optimized sampling strategy for the DSSCP.

MATERIALS AND METHODS

The DSSCP

The ongoing DSSCP has been based on the testing of meat-juice samples from finisher pig herds since 1995. The current sampling strategy has been in place since August 2001 and includes all herds with an annual kill of >200 slaughter pigs (representing 99% of all finisher herds in Denmark). The number of animals sampled at slaughter depends on herd size, with 60, 75, or 100 pigs sampled per herd per year. All samples are analysed at the Danish Institute for Food and Veterinary Research using the Danish mix-ELISA.

This test can detect O-antigens from at least 93% of all serovars known to be present in Danish pigs [17]. On the basis of testing, herds receive a monthly 'serological salmonella index' which is based on a weighted average of the results from the previous 3 months. The levels of index are low (index 1–39); medium (index 40–69); and high (index ≥ 70) [18]. In addition, testing is carried out in feed mills, multiplying, and breeding herds, slaughter plants, and on fresh pork.

There have been a number of changes to the DSSCP since its inception. These include penalizing farmers supplying pigs with an unacceptable level of serological salmonella index, lowering the positive cut-off for the Danish mix-ELISA, and introducing a fourth level of index (nil) where low risk herds are sampled less frequently [12]. The latter has been in place since July 2005.

The data

Two extracts of data were obtained from the central database of the DSSCP. These comprised the results of meat-juice testing conducted from 1 January 1995 to 31 May 2005 inclusive. The first data extract provided for each farm a unique identifier and details of farm location. The second data extract provided details of the 6992082 carcasses tested over the 10-year study period. Details included the unique farm identifier, the date of sampling, and the result of the Danish-mix ELISA. For these analyses an ELISA optical density percentage (OD%) >20 was classified as positive. This is equivalent to an adjusted OD% of >10: the cut-off for positivity that has been used by the DSSCP since 1 August 2001 [18].

For the period 8 May 2002 until 30 September 2004 inclusive, herd-size data were extracted from the Danish Central Husbandry Register and health status data were extracted from the Danish Specific Pathogen Free (SPF) Company.

Statistical analyses

Time-series analyses

We used time-series methods to describe the components of trend, cyclicity, and autocorrelation in the data. We reasoned that the exploration of trends in the data would facilitate the evaluation of the control programme, and that identification of seasonality or other temporal autocorrelation has the potential to inform surveillance.

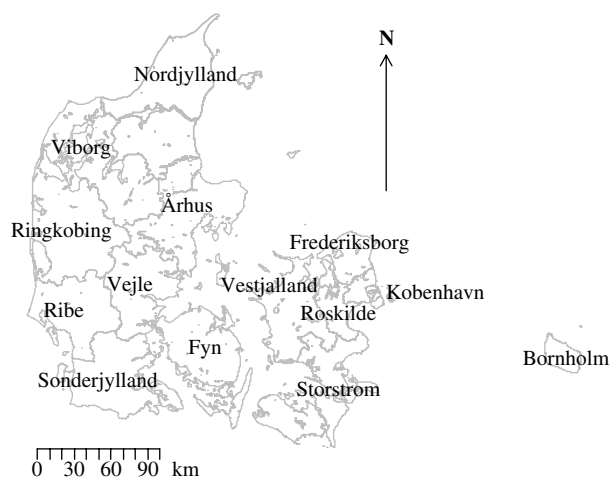


Fig. 1. Map of Denmark showing location of regions. The central island group containing Vestjylland, Roskilde, Storstrøm, København and Frederiksborg is Zealand. The largest land mass is Jutland.

To describe the trend, the weekly incidence risk (IR) of seropositivity was presented as a time-series graph. The presence of trend was formally tested using a bootstrapped Spearman test [19]. The positive results were then stratified into three levels: low (10–20 adjusted OD%); medium (21–50 adjusted OD%); and high positive (>50 adjusted OD%) and the weekly IR of the three strata were plotted. For all time-series plots loess smoothing splines were applied to the raw time series to emphasize the major features while reducing distraction from random variation [20].

The time series was stratified by region to investigate spatial variation in temporal patterns. Each regional series was further stratified by level of positivity. Because there were very few farms in the København region ($n=8$), they were aggregated with farms from a larger adjacent region, Frederiksborg. A regional map of Denmark is shown in Figure 1.

The data were rendered stationary to provide a degree of replication within the series facilitating further statistical analysis [21]. We first took the log of the proportion of samples positive to stabilize the variance and then detrended the series by fitting *a priori* a second-order polynomial to stabilize the mean. These transformations were performed on the overall time series and then for each region individually, allowing a more detailed examination of spatio-temporal variation. The model residuals were then examined as a near-stationary time series.

To identify the presence of seasonal effects we first examined grouped box plots and seasonal sub-series

plots. Second, we tested the statistical significance of calendar month by fitting an autoregressive linear model and calculating the coefficient of determination (R^2_{autoreg}) using the method proposed by Moineddin *et al.* [22]. Third, to identify cyclicity including that produced by seasonality, a periodogram was plotted [21]. This involved fitting sinusoidal waves with a discrete set of frequencies (Fourier frequencies) to the data. Using this technique the data were transformed to reveal cyclical behaviour. The practical value of the periodogram is that it can identify frequencies which are not always predictable before data are examined.

Temporal autocorrelation in the aggregated weekly data was identified using lagged scatter-plots, autocorrelation (ACF) plots and partial autocorrelation (PACF) plots. An autoregressive moving-average (ARMA) process was fitted to the data with examination of ACF and PACF plots used to identify the starting orders of the process [23]. Model parameters were estimated by maximum-likelihood methods using the Kalman filter. Overall model fit was assessed using Akaike Information Criteria [21]. Residuals were examined in three ways: plotting as a time series; checking the autocorrelation function; and applying the Ljung–Box test for independence [24]. Model performance was estimated by calculating the amount of the original series variance that was explained by the model [25]. This is analogous to the coefficient of determination (R^2) of a linear model.

Predictive modelling

For forecasting two subsets of the series were taken: (1) from weeks 1–487 to construct the model; and (2) from week 488 onwards for validation. We applied log transformation and differencing of the raw series to achieve stationarity. For forecasting we did not apply a polynomial fit since it places global assumptions on the data which may poorly estimate the fit beyond the range of the period of interest [21].

Identification of the order of differencing was by examination of ACF plots and estimating the variance of the series. An autoregressive integrated moving-average (ARIMA) process was fitted using the methods and diagnostics described previously. The predictive ability of the model was evaluated by two methods: (1) plotting the forecast and its 90% tolerance intervals and comparing it with the observed data; and (2) calculating the root-mean-squared percentage error criterion [26].

Farm-level investigation

For this part of the study we used data from 8 May 2002 to 30 September 2004. This was for two reasons: (1) for this 127-week period we had additional farm-level variables for 86% of farms allowing a stratified analysis to be conducted, and (2) computational constraints.

Repeated measures on the same farm would be expected to be correlated and determining the correlation structure has the potential to inform sampling strategy. For every week that a particular farm was sampled the outcome was defined as positive if at least one result for the week was positive and negative if all results for the week were negative. There were 11 754 farms contributing 697 877 farm weeks which were approximately uniformly distributed throughout the period of interest.

For stratification a subset comprising 86% of the original 11 754 farms for which farm-level variables were available was extracted. This comprised 10 064 farms and 609 537 farm-weeks. These farms were stratified into two herd-size levels: small (<700 slaughter pigs, $n=5605$) and large (≥ 700 slaughter pigs, $n=4459$). A second stratification was made by health status: conventional ($n=7028$) and SPF ($n=3036$). We defined conventional herds as those not in the Danish SPF system. The proportion of positive farm-weeks for the period was expressed as the IR.

The serial correlation of these repeated binary outcomes was explored using the lorelogram [27]. This is the mean log odds ratio between observations at each weekly time lag. It is a counterpart to the variogram approach for continuous longitudinal responses [28]. Both allow for the irregular spacing between sampling times on different farms and approximate an 'average' measure of temporal dependency of the repeated measures on farms. We produced lorelograms for all farms, and then for the different strata to identify if herd size and health status influenced the temporal dependency structure.

RESULTS**Time-series analyses**

Figure 2 presents a loess smoothed plot of all positive results over the study period, stratified by levels of positivity. With regard to all positive results, the first 3 years of the control programme had the highest IR (10–16%). This was followed by a decline in IR from

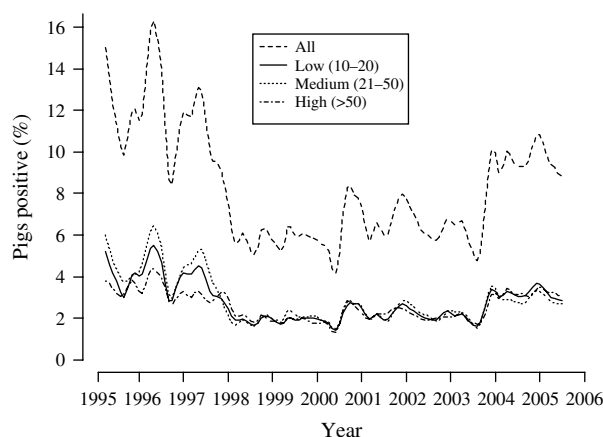


Fig. 2. Loess smoothed plot of the percentage of pigs positive for the Danish mix-ELISA stratified by level of positivity.

mid-1997. From 1999 to 2002 there was a period of relative stability with the percentage positive between 4% and 8%. In mid-2003 there was a large rise in positivity from 5% to 10% over 1 month. The IR remained at about 10% until the end of the study period. The Spearman test gave a value of -0.28 ($P < 0.001$) confirming the overall decreasing trend. The variance in the series was higher with higher IRs: the variance from 1995 to 1997 was 7.0, from 1998 to 2002 it was 1.4, and from 2003 onwards it was 3.9. In addition, in this first period (1995–1997) there is an apparent seasonal pattern to the data with visible peaks in February and November.

These trends were broadly followed for all three strata, although in the first 3 years of the DSSCP the series for the high positive strata (>50 adjusted OD%), was consistently lower than the other strata and, in the period from late 1996 to late 1997, did not show the peak shown by the other strata. The heterogeneity seen in the time series for high positive strata was accentuated when these were stratified by region. There was considerable variation across the country (Fig. 3*a–d*). However, even in the regions in Zealand (Fig. 3*d*) and on Bornholm (Fig. 3*c*, solid line), where the level and variation in positivity was generally lower, the overall trend showed an initial fall in positivity until 1998, then a plateau until a rise in 2003. Figure 3*a* (regions in the north of Jutland) shows that the time series for Århus has a different pattern compared with the other regions in the area.

When the series was stratified by region alone there were time periods when it appeared as if all regions were simultaneously experiencing a rise in the

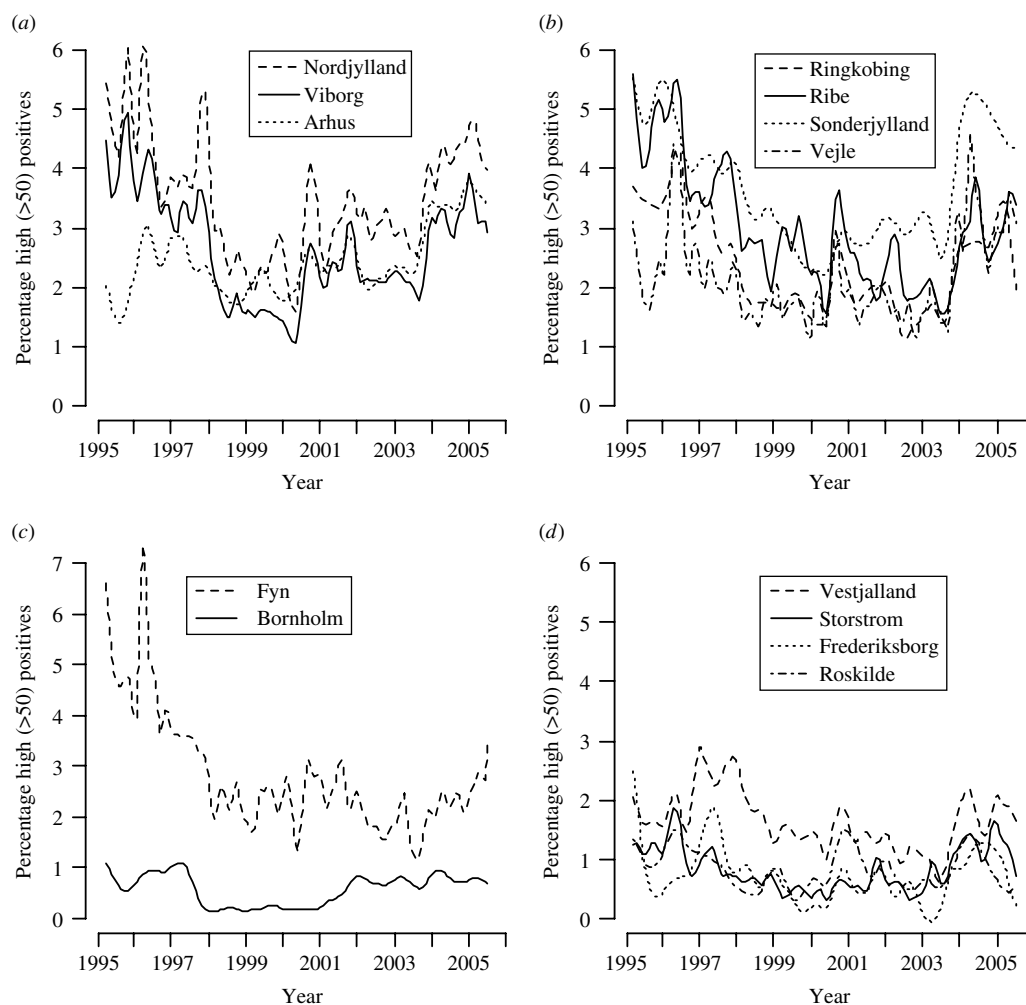


Fig. 3 (a–d). Loess smoothed plots of the percentage of pigs in the high positive strata (adjusted OD % > 50) for the Danish mix-ELISA stratified by region.

percentage of pigs positive. This was most noticeable in early 1996, early 1997, and late 2003 (not shown). Moreover, there were clear regional differences in IR: regions in the north and south of Jutland, and in the west of the country had higher IRs compared with the east.

Figure 4 presents the model residuals for the series once detrended by taking the log and fitting a second-order polynomial. The results for Denmark and for the main pig-producing counties are broadly similar. All show peaks of positive residuals in early 1996, early 1997, and late 2003 as in the raw data. However, a prominent peak of positive residuals in late 2000 only occurred in Nordjylland, Viborg, Ringkøbing, and Ribe (circles in Fig. 4*b, c, e, f*). These four counties are contiguous in the north-west of Jutland. This peak was less evident in Århus, Vejle and Fyn and absent in Sønderjylland and on Zealand and Bornholm.

There was large variability between years and no indication of cyclicity in the grouped box and seasonal sub-series plots (not shown). The R^2_{autoreg} for the autoregressive linear model was <1% indicating that calendar month was a very poor predictor of the series. The periodogram of the complete time series ($n=544$ weeks) had the most prominent peak at a frequency of 3: this equates to a 181-week cycle (not shown).

Lagged scatter-plots indicated serial dependence up until at least 16 weeks. The autocorrelation plot showed strong autocorrelation but no indication of a regular pattern as would be expected if seasonality were a factor (not shown). Examination of the ACF and PACF plots indicated that an autoregressive (AR) model, with order >1, may be appropriate. Model fitting resulted in a best fit of AR(3) to the stationary series. The parameter estimates were 0.63 [95% confidence interval (CI) 0.55–0.71], 0.06 (95%

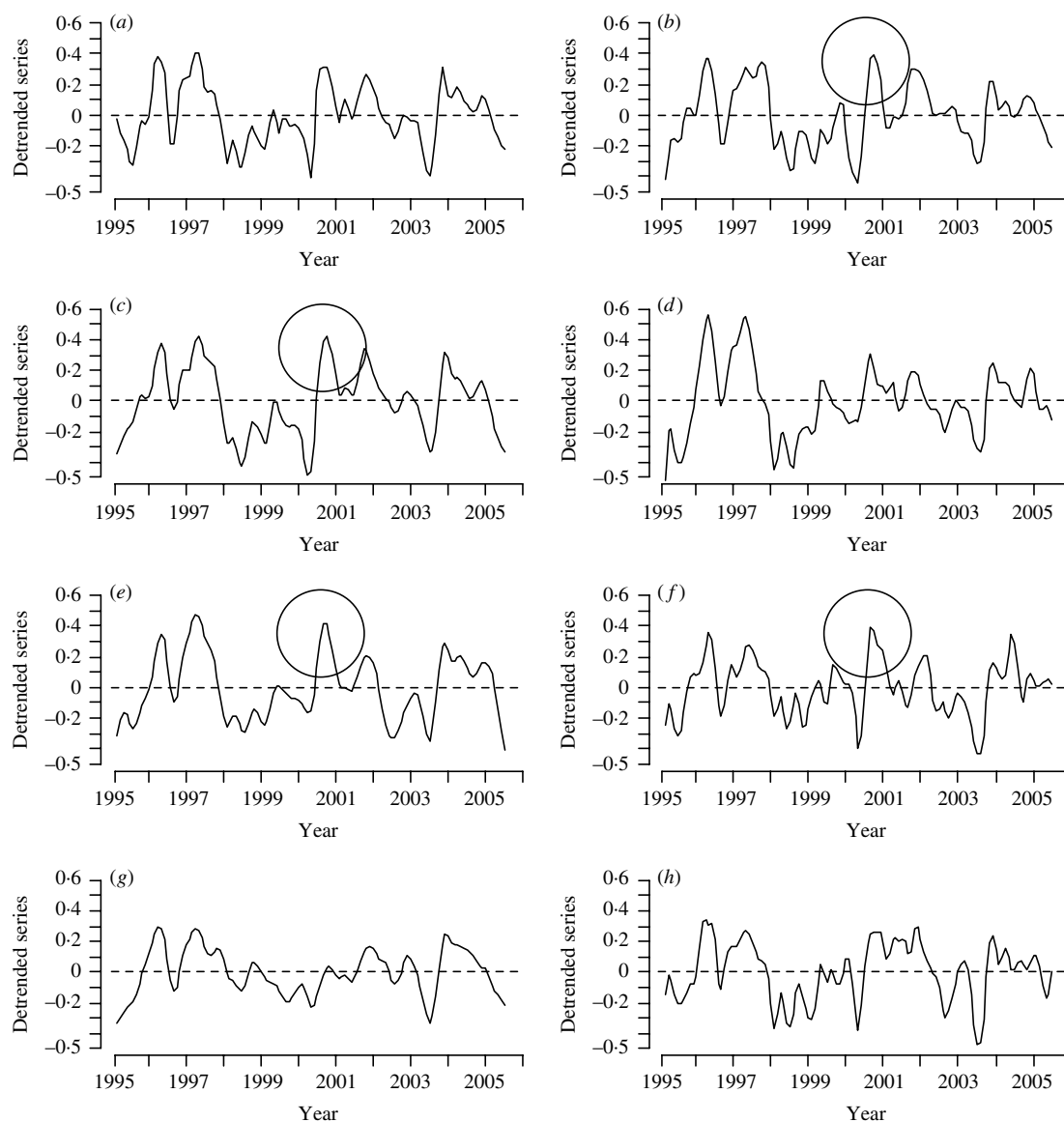


Fig. 4 (a–h). Plot of stationary time series for (a) Denmark, (b) Nordjylland, (c) Viborg, (d) Århus, (e) Ringkøbing, (f) Ribe, (g) Sønderjylland, and (h) Fyn respectively. The raw series has been logged to stabilize the variance and then detrended with a second-order polynomial. The dashed horizontal line is at the median percentage positive (0). The circles highlight the peak of positive residuals seen in counties in the north-west regions of Jutland in late 2000.

CI -0.04 – 0.16 , 0.19 (95% CI 0.10 – 0.27) for AR(1), AR(2) and AR(3) respectively. The residuals were randomly distributed in time. They were not significantly autocorrelated when tested by the ACF plot and the Ljung–Box test. The amount of the stationary series variance that was explained by fitting the model was 69%.

Predictive modelling

For forecasting an ARIMA (0, 1, 2) model gave the best fit. The parameter estimates for subset (i) were

-0.34 (95% CI -0.43 to -0.25) and -0.14 (95% CI -0.23 to -0.05) for MA(1) and MA(2) respectively. The residuals were not significantly autocorrelated and the amount of the logged series variance that was explained by fitting the model was 86%.

When the forecasts were superimposed on the logged observed data the observed value lay well within the 90% tolerance limits of the forecast (not shown). The root-mean-squared percentage error criterion for the 1-week-ahead forecast was 8.4% indicating the model had a reasonable forecasting ability.

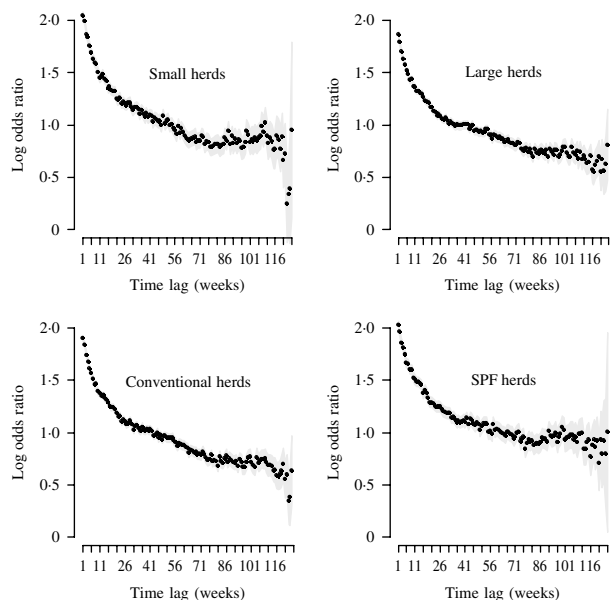


Fig. 5. Lorelograms (estimated mean log odds ratio as a function of lag time) for stratified data from May 2002 until September 2004. Grey shading is the 95% confidence intervals around the log odds ratios.

Farm-level investigation

The farm-week IR (the proportion of positive farm-weeks) for all 11 754 farms was 10.6% (95% CI 10.4–10.9). For the subset of 10 064 farms there were statistically significant differences between the strata: small and large farm IRs were 9.3% (95% CI 9.0–9.7) and 12.2% (95% CI 11.8–12.6) respectively. Conventional and SPF farm IRs were 11.3% (95% CI 10.9–11.6) and 10.1% (95% CI 9.6%–10.5) respectively.

The lorelograms for the stratified data are shown in Figure 5. There were two points of interest that apply to all the plots, and the equivalent plot for all 11 754 farms (not shown). First, the log odds ratios decrease rapidly up to a lag of 10 weeks, beyond which the decrease is more gradual. At lag 1 the odds ratio is approximately 7 indicating that positive farm-weeks tend to follow each other. These first 10 weeks show statistically significant temporal dependency, but beyond that there is still a strong downward trend in the lorelogram with the mean stabilizing at about 66 weeks. Second, the mean remains at levels well above zero suggesting that positive correlation remained at large time separations. This occurs as a result of heterogeneity between farms: some farms having relatively frequent positive weeks and some having very few or none.

There were also subtle differences in the plots for different strata. The two strata with the lowest IR (small and SPF herds) had mean odds ratios of eight at lag 1, and stabilized at week 56 with a mean odds ratio of 2.5. The two strata with the highest IR (large and conventional herds) had mean odds ratios of 6.5 at lag 1, and stabilized at week 76 with a mean odds ratio of 2.

DISCUSSION

This study has used accumulated data from the first 10 years of the DSSCP to investigate key temporal epidemiological features, explore the potential for forecasting and address the question of temporal redundancy. Our focus has been on the application of both established and novel techniques to respond to the need for optimization in the DSSCP.

The raw series showed a significant decreasing trend which was complex. The dramatic reduction in IR over the first 3½ years was predominant and probably due to the effect of the control programme [11]. From mid-1998 there was relative stability until a sudden rise in September 2003 when the IR rose from 6% to 10%. This rise was so sudden that a laboratory error was suspected and it was subsequently found that a problem in August 2002 had resulted in under-reporting of positive results [29]. This increase in seroprevalence may have been a factor in the small rise in pork-derived human salmonellosis cases in 2003, compared with 2002. However, 2004 saw a continuation in the long-term trend of a reduction in human cases that had been evident since 1996 [3].

We decided to stratify by level of positivity because differences exist in the pigs' serological response depending on the serovars involved, time since infection, and infection pressure within a herd [30]. Most serovars other than *Salmonella* Typhimurium give only a moderate serological response with the Danish Mix-ELISA [31]. Therefore, our finding of a different trend for the high positive strata (>50 adjusted OD%), in the first 3 years of the DSSCP, may indicate that at that time there was relatively greater contribution from serovars other than Typhimurium (Fig. 2). This could explain why the time series for the high positive strata in Århus shows a different pattern compared with the other regions in the north of Jutland (Fig. 3a). The pattern of a low, almost stationary percentage of pigs in the high positive strata seen in the eastern counties (Fig. 3d) may be a result of either infection with serovars with low test

sensitivity, or low herd-level infection pressures, or both.

Regional stratification identified times (early 1996 and early 1997) when the whole country was experiencing an increased IR. This may point to a seasonal pattern or to laboratory anomalies. The regional differences in IR are in agreement with earlier work [9, 17] and may be associated with large herd size: regions with high IR have the largest pig populations and larger pig farms [32]. However, the relationship between herd size and *Salmonella* seroprevalence is complex, with some studies identifying a positive association between herd size and seroprevalence [9] and others identifying a negative association [33].

The aggregations of positive residuals in the north-west regions of Jutland in late 2000 suggest a local epidemic. This could be due to the dissemination of contaminated feed or infected pigs throughout the area, possibly in combination with a regional practice that would favour rapid spread such as sharing of seasonal workers or machinery.

We found no evidence of a seasonal pattern in seroprevalence in our analysis. This is in agreement with a European-wide longitudinal study of seroprevalence in finisher pigs from October 1996 and May 1999 [34]. Similarly, temporal studies of passive laboratory-based surveillance data from 1991 to 2001 in Ontario [35] and Alberta, Canada [36] found no seasonal pattern in *Salmonella* isolation from production animals, including pigs. Pigs are likely to be infected with *Salmonella* from a variety of sources including herd mates, introduced animals, contaminated feed, rodents, and the environment. None of these need be seasonally distributed.

Previous work on the DSSCP reported a seasonal pattern of seroprevalence throughout the period 1995–1997 [9–11]. This was characterized by a summer trough and a late winter–early autumn peak. We also found visible peaks in February and November from 1995 to 1997 which may be due to seasonal effects, laboratory error or simply random variation. It would appear that, if inspected over a long enough period, the distribution of *Salmonella* seroprevalence does not follow a consistent seasonal pattern. Thus, we conclude that there is no apparent benefit in targeting sampling to particular times of the year. The seasonal pattern in human cases is much clearer, with consistent reports of a late summer–early autumn peak [11, 37].

Our finding of a peak in the periodogram analysis equating to a cycle of 181 weeks probably reflects the

residual peaks that are visible in Figure 4 in late 1996, 2000 and late 2003. This apparent cyclicity is probably artefactual and due to laboratory problems in late 2003 [29] and in late 1996 (P. Willeberg, personal communication, 2005). The peak in 2000 was due to a regional excess of *Salmonella* seropositivity in the north-west of Jutland.

Fitting the AR process to the data provided a temporal summary for the logged polynomial-detrended series. A second-order polynomial was fitted *a priori* allowing us to visualize and explain the complex trend in the national time series in three stages: (1) the dramatic effect of a successful control program (1995–1998); (2) a period of relative stability (1999–2002); and (3) a period of increasing national seroprevalence coincident with the laboratory problem in September 2003. We found that the current value of the percentage of pigs positive for the Danish-mix ELISA in any given week was dependent on the three preceding values, with most weight on the immediately preceding value. Our findings are identical to the relative weighting of 3:1:1 which is currently used to calculate the index for identifying herds for intervention in the DSSCP [18].

Forecasting is inherently challenging as it involves making assumptions based on past behaviour in the face of random variation. We used an ARIMA model that analyses the series as a function of its past values (AR), its trend (I) and its abrupt changes in the near past (MA) [23]. As it does not require a stationary series at the outset the potential problem of imposing global assumptions on data that may poorly estimate the fit beyond the range of the period of interest are overcome [21]. ARIMA models have traditionally been applied in the financial sector but are increasingly being used in medicine; recently as a tool for anomaly detection public health surveillance [38]. Forecasting nationally 1-week-ahead as done here would provide a baseline to which observed data could be compared. Observations outside a preset alarm threshold could signal an investigation to ascertain if the problem was geographically widespread or clustered within particular regions. Forecasts themselves could be done on a regional basis allowing for regional differences in seroprevalence. Within this context we envisage real-time calibration of the model, as additional data fed into the model should improve its predictive ability.

The lorelogram has been used to explore serial correlation of repeated binary outcomes for constipation treatment efficacy [39] and schizophrenia

symptoms [27]. Although we are not aware of the previous use of the lorelogram for optimizing a sampling strategy, the temporal variogram (a counterpart to the lorelogram for a continuous response variable), has been used for this purpose [14–16]. Our finding of a 10-week period of statistically significant temporal dependency indicates that at the farm level there is a strong ‘temporal memory’ up to (and to a lesser extent beyond), that lag. Extending this idea further would suggest that there is little value in sampling more frequently than every 10 weeks on the average farm. However, this theoretical approach needs to be balanced against the practicalities and advantages of the current continuous sampling and intervention strategy. Nonetheless, sampling at this reduced frequency (every 10 weeks) might be investigated for those herds enrolled in the ‘risk based’ scheme which has been running since July 2005 [12]. This scheme requires one sample per month to be taken from herds with a *Salmonella* index level of nil and a minimum 10 negative meat-juice samples in the last 6 months. To date 50% of herds meet these criteria.

It appears that there is a stronger ‘episodic’ effect (i.e. faster decay in the lorelogram before reaching relative stability) in small and SPF herds when compared with large and conventionally managed herds. The more persistent temporal dependency in the large and conventional herds indicates that these herds could benefit most from a reduced frequency of sampling. However, the results for the different strata must be interpreted with caution as there is potential for bias here. Our choice of binary outcome (i.e. a farm-week is positive if it has at least one result positive) would predispose large herds to having more positive farm-weeks than small herds purely by chance as they are sampled more frequently. The log odds ratios between observations at very long lag periods should also be interpreted with caution as there are few pairs contributing to this lag.

Due to the near complete coverage of sampling in the DSSCP (all herds producing >100 finishers per annum prior to 1 August 2001, all producing >200 pigs at or after 1 August 2001) our dataset is effectively a census of the population of Danish finisher herds. This has enabled inferences to be made about the *Salmonella* status of all Danish finisher herds over the period January 1995 to May 2005. However, the sampling scheme used prior to August 2001 resulted in pigs from large herds being proportionally over-represented. The results from this early period may be subject to this ascertainment bias.

We have applied time series and longitudinal analytical methods to identify patterns in both national and farm-level data for sub-clinical salmonellosis in Danish swine. These findings have direct and practical applications for both farm-level sampling strategies and national-level aberration detection which potentially could result in a more cost-effective surveillance strategy.

ACKNOWLEDGEMENTS

We acknowledge Bodil Ydesen (Danish Meat Association) for accessing the herd size and health status data.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Nielsen B, Wegener HC.** Public health and pork and pork products: regional perspectives of Denmark. *Revue Scientifique et Technique de l'Office International des Epizooties* 1997; **16**: 513–524.
2. **Hald T, et al.** A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Analysis* 2004; **24**: 255–269.
3. **Ministry of Family and Consumer Affairs.** Annual Report on Zoonoses in Denmark 2004. 2005. Copenhagen.
4. **Botteldoorn N, et al.** Salmonella on pig carcasses: positive pigs and cross contamination in the slaughterhouse. *Journal of Applied Microbiology* 2003; **95**: 891–903.
5. **Pearce RA, et al.** Studies to determine the critical control points in pork slaughter hazard analysis and critical control point systems. *International Journal of Food Microbiology* 2004; **90**: 331–339.
6. **Rostagno MH, et al.** Preslaughter holding environment in pork plants is highly contaminated with *Salmonella enterica*. *Applied and Environmental Microbiology* 2003; **69**: 4489–4494.
7. **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.
8. **Danske Slagterier.** Transport and pre-slaughter treatment (http://www.danskeslagterier.dk/smcms/Danish_English/Danish_Quality/2_transport/2_1_general/2_1_4_distances/Index.htm?ID=4118). Accessed 6 June 2006.
9. **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.

10. **Christensen J, Rudemo M.** Multiple change-point analysis applied to the monitoring of Salmonella prevalence in Danish pigs and pork. *Preventive Veterinary Medicine* 1998; **36**: 131–143.
11. **Hald T, Andersen JS.** Trends and seasonal variations in the occurrence of Salmonella in pigs, pork and humans in Denmark, 1995–2000. *Berliner und Munchener Tierarztliche Wochenschrift* 2001; **114**: 346–349.
12. **National Committee for Pig Production.** Annual Report 2005. 2006. Copenhagen.
13. **Stärk KD, et al.** Concepts for risk-based surveillance in the field of veterinary medicine and veterinary public health: Review of current approaches. *BMC Health Services Research* 2006; **6**: 20–28.
14. **Salvatori V, et al.** Estimating temporal independence of radio-telemetry data on animal activity. *Journal of Theoretical Biology* 1999; **198**: 567.
15. **Dowdall M, et al.** Geostatistical analysis as applied to two environmental radiometric time series. *Environmental Monitoring and Assessment* 2003; **83**: 1–16.
16. **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.
17. **Mousing J, et al.** Nation-wide Salmonella enterica surveillance and control in Danish slaughter swine herds. *Preventive Veterinary Medicine* 1997; **29**: 247–261.
18. **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.
19. **Henderson AR.** The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clinica Chimica Acta* 2005; **359**: 1–26.
20. **Bowman AW, Azzalini A.** *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press, 1997, pp. 193.
21. **Diggle P.** *Time Series: A Biostatistical Introduction*. Oxford: Oxford University Press, 1990, pp. 257.
22. **Moineddin R, et al.** Autoregression as a means of assessing the strength of seasonality in a time series. *Population Health Metrics* 2003; **1**: 10.
23. **Box GEP, Jenkins GM, Reinsel GC.** *Time Series Analysis: Forecasting and Control*, 3rd edn. Englewood Cliffs, NJ: Prentice Hall, 1994, pp. 598.
24. **Ljung GM, Box GEP.** On a measure of lack of fit in time series models. *Biometrika* 1978; **65**: 297–303.
25. **Lopez-Lozano JM, et al.** Modelling and forecasting antimicrobial resistance and its dynamic relationship to antimicrobial use: a time series analysis. *International Journal of Antimicrobial Agents* 2000; **14**: 21–31.
26. **De Gooijer JG, Hyndman RJ.** Twenty-five years of time series forecasting. *International Journal of Forecasting* 2006; **22**: 443–473.
27. **Heagerty P, Zeger S.** Lorelogram: a regression approach to exploring dependence in longitudinal categorical responses. *Journal of the American Statistical Association* 1998; **93**: 150–162.
28. **Diggle P, et al.** *The Analysis of Longitudinal Data*, 2nd edn. Oxford: Oxford University Press, 2002, pp. 379.
29. **Ministry of Food Agriculture and Fisheries.** Annual Report on Zoonoses in Denmark 2003. 2004. Copenhagen.
30. **Nielsen B, et al.** The serological response to Salmonella serovars typhimurium and infantis in experimentally infected pigs. The time course followed with an indirect anti-LPS ELISA and bacteriological examinations. *Veterinary Microbiology* 1995; **47**: 205–218.
31. **Lo Fo Wong DM, et al.** Recovery of Salmonella enterica from seropositive finishing pig herds. *Veterinary Microbiology* 2003; **97**: 201–214.
32. **Danske Slagterier.** Statistics (http://www.danskeslagterier.dk/smcms/Danske_Slagterier_UK). Accessed 7 March 2005.
33. **van der Wolf PJ, 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.
34. **Lo Fo Wong D, et al.** A European longitudinal study in Salmonella seronegative- and seropositive-classified finishing pig herds. *Epidemiology and Infection* 2004; **132**: 903–914.
35. **Zhang X, et al.** Detection of clusters of Salmonella in animals in Ontario from 1991 to 2001. *Canadian Veterinary Journal* 2005; **46**: 517–519, 522–513.
36. **Guerin MT, et al.** A temporal study of Salmonella serovars in animals in Alberta between 1990 and 2001. *Canadian Journal of Veterinary Research* 2005; **69**: 88–99.
37. **Guerin MT, Martin SW, Darlington GA.** Temporal clusters of Salmonella serovars in humans in Alberta, 1990–2001. *Canadian Journal of Public Health* 2005; **96**: 390–395.
38. **Le Strat Y.** Overview of temporal surveillance. In: Lawson AB, Kleinman K, eds. *Spatial and Syndromic Surveillance for Public Health*. Chichester: John Wiley & Sons Ltd, 2005, pp. 13–29.
39. **Choi L, et al.** Estimating treatment efficacy over time: a logistic regression model for binary longitudinal outcomes. *Statistics in Medicine* 2005; **24**: 2789–2805.