ELSEVIER

Preventive Veterinary Medicine 34 (1998) 191–203

PREVENTIVE VETERINARY MEDICINE

Herd size and sero-prevalence of *Salmonella enterica* in Danish swine herds: a random-effects model for register data

Bendix Carstensen *, Jette Christensen

Danish Veterinary Laboratory, Bülowsvej 27, DK-1790 Copenhagen V, Denmark

Accepted 18 July 1997

Abstract

The association between herd size and sero-prevalence of *Salmonella* was assessed in a random-effects model with herd size, county and date of slaughter as fixed effects. A total of 510 915 meat-juice samples from 14 593 herds located in 13 counties in Denmark was included in the study. A random-effects model was developed from separate models for smaller strata of data from herds with approximately equal sizes. The combined model was analysed and the results reported. Herd size was positively associated with the sero-prevalence of *Salmonella enterica*, but the size of the association was biologically of little importance, because the within-herd and the between-herd variations were relatively large in comparison. The relative magnitudes of the variance components indicated that factors associated with both the herd level and the pig level could be important in the prediction of seroprevalence of *S. enterica*. © 1998 Elsevier Science B.V.

Keywords: Pig—microbiological diseases; *Salmonella enterica*; Herd size; Random effects; Modelling

1. Introduction

Since January 1995, a nationwide *Salmonella enterica* surveillance programme in swine herds has been operating in Denmark. The programme was started out of consumer concern (Baggesen et al., 1996), and is described by Mousing et al. (1997), and Nielsen et al. (1995).

In short, all swine herds with an expected yearly kill exceeding 100 pigs are monitored serologically by a MIX-ELISA test (Nielsen et al., 1995) of meat-juice from samples taken at slaughter. Herds with a moderate proportion of pigs with antibodies

* Corresponding author. E-mail: bxc@svs.dk

PII S0167-5877(97)00072-X

against *Salmonella* are required to receive veterinary advice to control *Salmonella*, and herds with a higher proportion are additionally required to have their pigs slaughtered under special hygiene precautions.

Herds producing more than 2000 pigs per year (11.4% of the herds) produced 56.1% of the pigs slaughtered in Denmark during 1995 (Danske Slagterier, 1996) and the trend is that the herd size is increasing. Since more and more pork will originate from large herds, it is important to evaluate whether the occurrence of *Salmonella* increases with increasing herd size.

There are few published studies on the effect of herd size on the occurrence of *Salmonella* in swine. Preliminary analyses have indicated increased risk of *Salmonella* infections in slaughter pigs with increasing herd size (Baggesen et al., 1996; Mousing et al., 1997) and in cattle the highest incidence of salmonellosis was found in the largest herds, with drained pen type (Kristiansen et al., 1985).

The effect of herd size on diarrhoea in swine herds is not clear and few investigations have been reported. The producer-recorded morbidity of diarrhoea in piglets 4–14 days of age was greater in large herds than small herds (Dewey et al., 1995). In a prospective study including 85 breeding units in East Anglia which experienced primary outbreaks of transmissible gastroenteritis (TGE) between December 1980 and October 1982, the incidence of recrudescence of TGE was higher in large herds (Pritchard, 1987). A Danish study on post-weaning diarrhoea (Svensmark et al., 1989) found that, adjusted for age at weaning, the incidence decreased with increasing herd size.

This paper will evaluate the effect of herd size on the sero-prevalence of *S. enterica* based on data from the surveillance programme. Since there are multiple measurements from each herd, we need to account for the dependence between samples from the same herd. Further, we wanted to evaluate the effects of date of slaughter (season) and geographical location of the herd. This led us to construct a random-effects model with random effects of herd, herd size by location interaction and herd size by date of slaughter interaction. In this model, we estimated fixed effects of geographical location (county) and date of slaughter as well as herd size.

A random-effects model allows interpretation of the size of the fixed effects in relation to the sizes of the variances of the random effects-that is, to quantify the effect of the known variables (included as fixed effects) in relation to the variation in those not known (modelled as random effects). This means that the focus of the statistical modelling will be moved from significance testing of effects to estimation of effects and interpretation of the importance of these relative to each other and relative to the subject matter.

2. Material

2.1. The surveillance programme

The data for this study were extracted from the Central Zoonosis Register (ZOOR) established as part of the nationwide *S. enterica* surveillance and control programme in Danish swine herds (Mousing et al., 1997). ZOOR is an official database owned by the Ministry of Food, Agriculture and Fishery but run in close cooperation among the

Danish Veterinary Service (DVS), the Danish Veterinary Laboratory (DVL), the Federation of Danish Pig Producers and Slaughterhouses (FDPPS), and individual slaughterhouses.

The role of ZOOR is to integrate data from the private and official institutions involved in the surveillance programme and to allow automatic selection of herds and carcasses to be sampled for serological testing. The data include: (1) identification of the farms from the Central Husbandry Registration Register (CHR-register) including geographical location; (2) monthly data on the number of finisher pigs slaughtered per farm from the central database of the cooperative slaughterhouses; and (3) serological and bacteriological test results from DVL.

In ZOOR, a herd is defined as the population of pigs raised at one farm. All herds with a production exceeding 25 pigs during the last 13 weeks are included in the surveillance programme. The sampling fraction for each herd varies by the size of the production, ranging from 11% of pigs slaughtered in herds producing less than 200 pigs per year to 3.3% in herds producing more than 2000 pigs per year (Mousing et al., 1997).

A procedure for automatic selection of carcasses for serological testing is used, and a meat sample of approximately 10 g (from the diaphragm, the tender loin, or the sternomastoideus muscle) is excised, frozen and sent to the DVL for serological examination of meat-juice (obtained by thawing) with a MIX-ELISA (Nielsen et al., 1995). The obtained serological test results are used as basis for automatical assignment of intervention level for the individual herds.

2.2. The analysis dataset

All test results from meat-juice samples in the surveillance programme taken between October 1994 and December 1995 were selected ($n = 538\,206$). Herds with missing data on county or production level were excluded, as were organic herds, free-range herds and multiplying herds. Samples with insufficient identification or samples producing insufficient meat-juice for examination were excluded. A total of 510 915 meat-juice samples (94.9%) from 14 593 herds located in 13 counties (out of 16 in Denmark) were included in this study.

The data on each sample comprised information on county, the annual number of pigs slaughtered (farm-level variables), date of sampling and OD-value (Optical Density)(sample-level variables). The OD-value used throughout is a transformed OD%: all OD% ≤ 11 are recorded as 1 and all other as the OD% minus 10. Thus, we have a measurement that is left-censored at 1.

The overall *S. enterica* sero-prevalence in the material was 5.4% using a cut-off OD-value 30, and 11.1% using a cut-off OD-value 10.

3. Statistical methods

3.1. Initial modelling at herd-level

In order to explore the data initially, we reduced the data to the average OD-value (of the non-censored values; i.e. of the OD-values > 1) per herd. This was plotted against herd size, and smoothed averages of these were calculated by three different algorithms.

Analysis of the herd averages circumvents the problems arising from the anticipated correlation between observations from the same herd, and could also be extended to take geographical variation into account—but prevents the use of information on date of slaughter, since the latter variable varies *within* herds.

3.2. Model for the complete material

The complete material of 510,915 individual samples was therefore analyzed with the scope of predicting the level of the serological response in samples of meat-juice, and how this varies with county, date of slaughter and herd size. The point is, that even if the herd-size effect is the focus, it is necessary to include the other effects too, in order to control possible confounding and in order to assess the relative sizes of the factors influencing the level of serological response.

In the following, we first describe the structure of the model, and the principles underlying the estimation procedure in the context of a model for normally-distributed data. Subsequently, we describe the extension of the approach to the skewed and heavily-censored data that we analyze.

A natural model to choose for the present material is a simple two-level random-effects model with a random between-herd effect, a combined within-herd variation and measurement error, and fixed effects of county, date of slaughter and herd size.

Thus, for a response y_{hi} from herd h and individual i (for the time being, think of it as the serological response):

$$y_{hi} = \beta_c + \kappa(d) + \phi(s) + a_h + e_{hi}. \quad (1)$$

In this model, the fixed effects are: β_c—the effect of county; $\kappa(d)$—the effect of date of slaughter; $\phi(s)$—the herd-size effect; and the random effects a_h—the herd-effect and e_{hi}—the combined within-herd variation and measurement error. The random effects are assumed to be independent and normally distributed with variances τ^2 and σ^2, respectively.

This model assumes that there is no interaction between herd size and the other variables and that the variance of the random effects is the same regardless of herd size. Further, we have left the precise functional form of the date effect ($\kappa(d)$) and herd size ($\phi(s)$) unspecified, so in order to make the model useful, we need to specify these two functions.

The date effect was modelled by cubic splines, i.e. in each of 7 intervals we fitted a cubic function of date, and constrained these to fit smoothly together at the joining points, which were chosen to be 25, 75 and 125 days on either side of 1 July 1995.

In order to model the herd-size effect, we adopted a model stratified by herd size in 150 strata, i.e. with some 100 herds (of approximately equal size) in each stratum. In each stratum we fitted model (1)—but now with the term $\phi(s)$ representing an intercept specific for the stratum with herd size s, and therefore absorbed in the county effect. In practical terms, this means that the dataset is split in 150 pieces (by s) and the following model is fitted in each:

$$y_{hi}^{(s)} = \beta_c^{(s)} + \kappa^{(s)}(d) + a_h^{(s)} + e_{hi}^{(s)} \quad (2)$$

where c indexes county (13 levels), and s stratum (i.e. herd size).

From model (2), we get 150 sets of estimates of county effects ($\beta_c^{(s)}$), 150 sets of date effects ($\kappa^{(s)}(d)$) and 150 sets of estimates of between and within herd variances (τ_s, σ_s), one set from each stratum. The 150 sets of fixed-effects estimates are not of interest per se, so a more reasonable model would be one where the *variation* of these were considered to be random, and only their average reported. This would be the model:

$$y_{hi} = \beta_c + \kappa(d) + \phi(s) + c_{cs} + b_{ds} + a_h + e_{hi} \quad (3)$$

where $\beta_c + \kappa(d) + \phi(s)$ represents the fixed effects of county, date of slaughter and herd size, respectively, and $c_{cs} + b_{ds}$ represents the random interaction effects. We assume that the random effects c_{cs}, b_{ds}, a_h and e_{hi} have variances υ^2, ω^2, τ_s^2 and σ_s^2, respectively. Note that we have assumed that there is no county × date interaction, and we do not specify how the variance components τ_s^2 and σ_s^2 depend on s.

The interpretation of model (3) requires some care, because the effect of county has to be evaluated against the variance of c_{cs} (υ^2); the effect of date against the variance of b_{ds} (ω^2); and the effect of herd size has to be evaluated against the sum of these two variances (since the change from one herd size to another involves change of two variance components).

3.3. Extension to a generalized linear model

The actual measurements y_{hi} are given as integers, ranging from 2 to over 100, with the values from 2 up to 40 appearing in decreasing frequency, while 1 represents all measurements equal to or lower than 1. Therefore, the data cannot be assumed to be normally distributed, and neither can any transformation bring them in a form that allows such an assumption to be reasonable.

Therefore, we adopted a model where not only the effect of the covariates but also the shape of the distribution is estimated. This is achieved through a model for the event $\{Y_{hi} \geq M\}$ for some value of M (in this case the cut-offs of 10 and 30 in the transformed OD-value) with ln(−ln)-link function, which is a generalized linear model.

This model can readily be extended to accommodate random effects as in the case of the general linear model used in the previous section. We simply take the *conditional* distribution of $\{Y \geq M\}$, conditional on the herd effect a_h and the measurement error, c_{hi}, to be binomial with:

$$-\ln\left[-\ln\left(\Pr\{y_{hi} \geq M | a_h, e_{hi}\}\right)\right] = \alpha_M + \beta_c + \kappa(d) + \phi(s) + c_{cs} + b_{ds} + a_h + e_{hi}. \quad (4)$$

An assumption of proportionality of hazards in the OD-distribution translates to the assumption that the covariate effects (and random effects) are the same on this scale, regardless of the choice of cutpoint M. (This is merely an assumption about the scale on which the covariates act). Now, all the arguments from the previous section on the normal case translate to this, only now everything takes place at the $-\ln(-\ln)$-scale.

The actual estimation was done as described by Wolfinger and O'Connell (1993), and implemented in the SAS-system macro %glimmix, available at distributions discs of SAS, and at ftp://ftp.sas.com/techsup/download/stat/ where macros suitable for pre- or post-6.12 version of SAS are available as glmm611.sas or glmm612.sas.

3.4. Practical estimation

Estimation in the model (4), can be accomplished in the following way:

(1) In each stratum, fit a generalised linear model that looks like Eq. (2):

$$-\ln\left[-\ln\left(Pr\{y_{hi} \geq M | a_h, e_{hi}\}\right)\right] = \alpha_M^{(s)} + \beta_c^{(s)} + \kappa^{(s)}(d) + a_h^{(s)} + e_{hi}^{(s)}.$$

This produces estimates of τ_s^2 and σ_s^2, and estimates of effects of county and date of slaughter for each stratum. We can thus arrange the estimates of the fixed effects in two, two-way tables classified by stratum on one side and county or date of slaughter, on the other.

(2) Fit a two way main-effects model to the estimates of county effects classified by county and stratum, using the inverse of the estimated variances as weights. This produces estimates of the county effects and of the effect of stratum (herd size), as well as a residual variance (i.e. the variance of the random interaction effect, υ^2).

(3) Fit a two way main-effects model to the estimates of date effects classified by date and stratum, using the inverse of the estimated variances as weights. This produces estimates of the date effect and of the effect of stratum (herd size), as well as a residual variance (i.e. the variance of the random interaction effect, ω^2).

(4) Combine the estimated effects of herd size obtained under 2 and 3 by adding them.

The county effect is parameterized by an intercept parameter for each county, but the date effect is parametrized by cubic splines in 7 intervals. Thus the procedure outlined above is straightforward for the county effects, but for the date effects it is necessary to use the estimates of the date effect at a number of dates. In this analysis, we used 25 equally spaced dates between 3 March 1995 and 29 October 1995 (10-day spacing).

The estimate of the date-stratum interaction is approximately invariant under different choices of the dates where the effect is evaluated (Appendix A).

3.5. Interpretation of results

Results from the model will be fixed-effect parameters (describing the effects of county, date of slaughter and herd size) and standard deviations of random effects (describing residual effects of unmeasured variables). Thus, for all effects it is necessary to consider the standard error of random measurement error (σ) and for all herd-level variables also the standard error of the random herd effect (τ) and on top of these also the standard error of the relevant interaction effects (υ^2 and/or ω^2).

Consider e.g. an estimated difference between two county parameters of 0.25. Since county is a herd-level variable this has to be evaluated against the size of the between-herd random effect (τ), the random effect of county $\times$ herd size (υ) and the random within-herd effect (σ) as well. If these three effects have standard errors of e.g. 0.4, 0.2 and 0.6, then the standard error of the joint effect is $\sqrt{0.4^2 + 0.2^2 + 0.6^2} = 0.75$. Thus, in such a case we would consider the county effect small. Note that this is only from the prediction point of view, i.e. if the county effect is important in prediction of levels of *Salmonella* sero-response. Whether there is significant difference in the ('mean') level of response between counties should be assessed by the estimated

variance–covariance matrix of the county effects, that results from the analysis (Step 2, in the description above).

The results from the analysis are shown as graphs of the fixed effects of county, date of slaughter and herd size (all with 95% pointwise confidence intervals), and a plot of the estimated between and within-herd standard deviations against herd size with a smoothed mean curve (Figs. 2 and 3). The latter also shows the size of the standard error of the random interaction terms.

4. Results

4.1. Simple herd-level analyses

From Fig. 1, it appears that there is a slight increase in average OD-level by herd size, but from the data points, it is also apparent that the variation between herds is enormous compared to the herd size effect. How this variation can be partitioned is discussed below.

4.2. Variance component model

The model is fitted separately for cut-offs at OD-value 10 (OD10) and OD-value 30 (OD30) to the dataset. The two analyses are not independent, but a proportional hazards

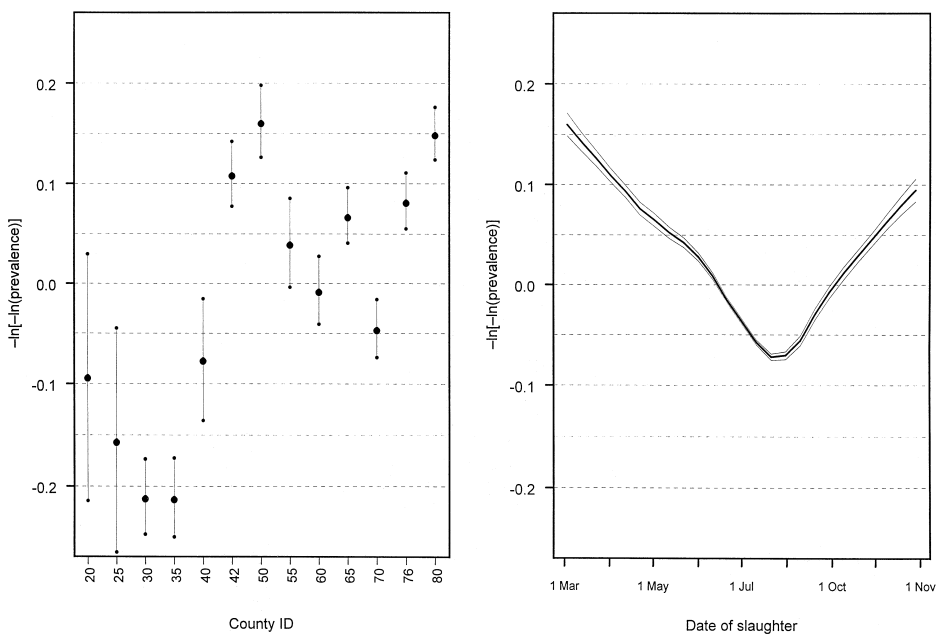

Fig. 1. Estimated effects of county and date of slaughter from the stratified random effects model for cut-off OD10, based on 510915 samples from 14593 herds. Effects of county and date are results from main effects models for the estimates from each of 150 strata defined by herd size. The thin lines indicate 95% confidence intervals for the mean.

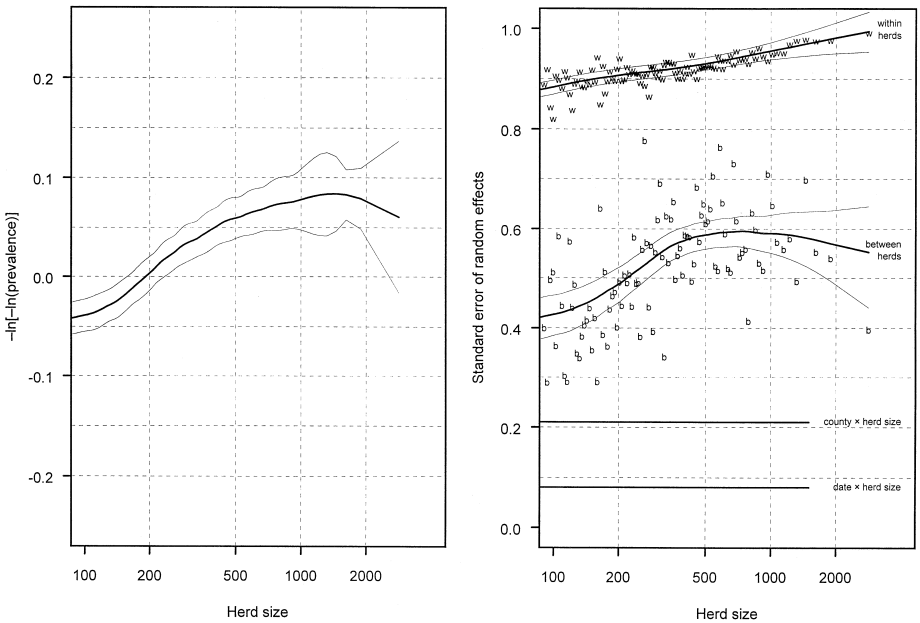

Fig. 2. Estimated effect of herd size and variations of random effects from the stratified random effects model for cut-off OD10, based on 510 915 samples from 14 593 herds. The effect of herd size is the result of adding the estimated herd size effect from the analyses of the county × herd size and date × herd size. The random interaction variations are residual variations from these analyses. The thin lines indicate 95% confidence intervals for the mean.

assumption means that they should yield identical results both for the fixed effects (except for the intercept) and random effects.

The results from the stratified variance component model for cut-off OD10 are given in Figs. 2 and 3 and results are summarized in Tables 1 and 2 for both cut-offs. As can be seen from the tables, the results for the two cut-points are fairly similar, hence only the results from OD 10 are presented as graphs.

The predominant picture is that the herd-size effect is the smallest of the fixed effects, although almost of the same order of magnitude as the other two fixed effects. The range of the fixed effect is around 0.2 on the $-\ln(-\ln)$-scale.

The within-herd variation (extra-binomial variation) is increasing with increasing herd size. This is partly attributable to the fact that the sample size per herd is larger the larger the herd is. The between-herd variation is roughly constant over herd sizes, suggesting that the factors unaccounted for at the herd level are not associated with herd size.

The size of the standard deviation of the between- and within-herd random effects is larger than the size of the fixed effects. The standard deviation of the county × herd size interaction is somewhat smaller, and the standard deviation of the date × herd size interaction, yet smaller again. In this context one should bear in mind that a 90% prediction interval for a random effect has a width of about three times the standard deviation ($2 \times 1.645 = 3.29$).

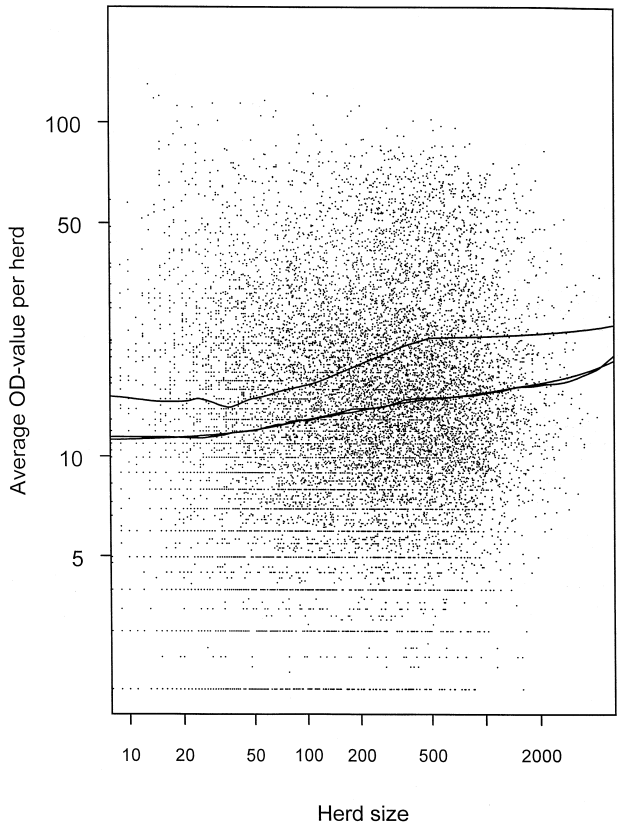

Fig. 3. Average OD-value per herd with smoothed averages by herd size. One point represents the average OD-value for pigs from that herd plotted against the size of the herd. The three lines represent results from three different smoothing algorithms.

Table 1
Estimated effects from the stratified random-effects model for *S. enterica* seroprevalence, for 510 915 meat juice samples from 14 593 swine herds in Denmark. All estimates are on the $-\ln(-\ln[p])$-scale

Parameter	Cut-off			
	OD10		OD30	
	Range	Median	Range	Median
Fixed effects (estimated effects)				
Herd size (200–2000 pigs/year)	0.00 – 0.08	0.08	−0.01 – 0.12	0.13
County (min–max)	−0.21 – 0.17	0.38	−0.22 – 0.16	0.38
Date (min–max)	−0.07 – 0.17	0.24	−0.05 – 0.12	0.17
Random effects (standard errors)				
Within herds (σ)	0.81 – 0.97	0.91	0.58 – 0.90	0.70
Between herds (τ)	0.28 – 0.78	0.51	0.35 – 0.80	0.58
County × herd size (υ)		0.21		0.26
Date × herd size (ω)		0.08		0.12

Table 2
Estimated effects on the prevalence scale (in % of samples sero-positive by a mix-ELISA for *S. enterica*) from the stratified random effects model, for 510 915 meat juice samples from 14 593 swine herds in Denmark. The estimated prevalences are for the indicated ranges of parameters from Table 1, but transformed to the prevalence scale. The baseline estimates are for herds with all other effects 0. The relative risks (RR) refer to ratios of prevalences between extremes of the ranges

Parameter	Cut-off			
	OD10		OD30	
	Prevalence	RR	Prevalence	RR
Fixed effects				
Baseline prevalence	13.5	–	1.55	–
Herd size (200–2000 pigs/yr)	13.5 – 15.7	1.16	1.49 – 2.48	1.66
County	8.5 – 18.5	2.18	0.56 – 2.87	5.13
Date	11.7 – 18.5	1.58	1.25 – 2.48	1.98
Random effects (± 1.64 × median)				
Within herds	0.0 – 63.4	–	0.00 – 26.78	–
Between herds	0.1 – 42.7	427	0.00 – 20.09	–
County × herd size	5.9 – 24.2	4.1	0.19 – 6.32	33.3
Date × herd size	10.2 – 17.3	1.7	0.62 – 3.37	5.3

Although the fixed effects are clearly significant, they are of little value in prediction of *Salmonella* levels. Any prediction will end up almost totally uninformative owing to the large random variations. This is clear from Table 2, where in particular the contributions from the between and within-herd effects render predictions uninformative.

5. Discussion

5.1. Statistical method

The statistical model chosen to describe the material has two aims: firstly to take the within-herd correlation into account and secondly to account for the residual variation by distributing it among different sources.

Apart from the herd effect, we have introduced two random effects: the date × herd size and the county × herd size interactions. The point in attributing the random variation to these sources instead of lumping it all together in an extra-binomial variation at some finer level (which would have to be at least as fine as a date × herd classification, requiring some grouping of date), is to explore potential factors responsible for the unexplained variation in the material.

As we find the within-herd variation to be the largest, some of the important factors are to be found at the within-herd-level. Also, since the between-herd variation is somewhat larger than the two interactions it may be inferred that other factors of interest

may be found at the herd level. The date × herd size interaction has a comparatively small standard error, indicating that the effect of date is not varying much between different herd sizes.

Since we are working with a binomial model, we will also find some extra-binomial variation. The way we are modelling this extra-binomial variation, it can be regarded as the imprecision in determination of the linear predictor for single units; this imprecision stems from the important yet unmeasured variables.

5.2. Biological interpretation

The serological measurement (OD-values) is predictive of the bacteriological status of the individual pig, but definition of sensitivity and specificity of an OD-value above e.g., 10, requires a definition of a condition of the individual pig that we aim to diagnose. This may be 'ever infected with *Salmonella*', 'carrier at time of slaughter', 'actually contaminating the slaughterhouse'. As long as this is not stated, sensitivity and specificity cannot be defined.

Strictly speaking, what we deal with is the risk that a given pig is recorded in the data-base with an OD-value above 10, respectively 30. So the results we report here are about risk factors for elevated serological response to *S. enterica*. This is of course closely connected to the actual *Salmonella* status of the herd (however that is defined), but to speak about sensitivity and specificity without specification of the condition that we are aiming to diagnose (in the individual pig) is not meaningful.

All investigated fixed effects were significant but that is simply due to the large sample size (510 915 meat-juice samples). The relative risk of exceeding OD30 associated with a ten-fold increase in herd size from 200 to 2000 pigs slaughtered per year was merely 1.7- of the same magnitude as the relative risk of a positive test between the date with the lowest prevalence (July 1995) and with the highest prevalence (November 1995). Similarly, the relative risk was 5.1 when the counties with the greatest difference in prevalence were compared.

Since the power of the study is high, it is important to evaluate the relative magnitudes of the estimates of the fixed effects and the magnitudes of the standard errors of the random effects. For example, the standard error of the herd effect (0.58) was approximately four times the range of the effect of herd size (-0.01 to $0.12 = 0.13$) on the $-\ln(-\ln)$-scale.

The interpretation is that there are other factors than herd size, county, and date that may be biologically more important because they potentially explain more of the variation in *Salmonella* occurrence.

The standard errors of the random effects depended on the herd size (Fig. 3) and the standard error of the variation between herds was slightly smaller than the standard error of the variation within herds (proportions of total variance are $0.58^2/[0.70^2 + 0.58^2 + 0.26^2 + 0.12^2] = 37\%$ and $0.70^2/[0.70^2 + 0.58^2 + 0.26^2 + 0.12^2] = 54\%$, respectively. Since both the variances within and between herds were relatively high, we conclude that the other factors may be acting at the herd level as well as at the pig level.

The factors acting at the pig level might be type of pen separation, pig density in the pens, other diseases, and distance to a pig excreting *Salmonella*. In contrast, factors

acting at the herd level might be wet feed/dry feed, on-farm produced feed, slurry/manure management, cleaning/disinfection procedures, and pig density in the geographical area. Some factors may act at different levels; for example all-in/all-out could be practised at the herd, house, or pen level.

The potential complexity of the factors influencing the occurrence of *Salmonella* antibodies indicates that a series of investigations is necessary to fully describe them. Some studies will have to focus on herd level factors while others should focus on the pig level.

5.3. Representativeness

There was no indication that the study sample should not be representative of the Danish swine population with regard to herd size, county, and date of slaughter after June 1995.

The study included $14\,593/16\,672 = 88\%$ of the farms with pigs above 50 kg live weight and 2.9% of the pigs produced in Denmark during 1995. The sampling intensity increased during 1995, and therefore more samples were taken in the second part of 1995 compared to the first part but the comparison July to November should be valid.

6. Conclusions

Herd size had a statistically significant effect on the sero-prevalence of *S. enterica*—but it was biologically of little importance, because the within-herd and the between-herd variances were relatively large in comparison. The relative magnitude of the variance components indicated that factors associated with both the herd level and the pig level could be important in the prediction of sero-prevalence of *S. enterica*.

Appendix A

The estimate of the date-stratum interaction is obtained by:

1. Evaluate the date-effect in each of the 150 strata, at a set of prespecified dates. This gives dataset of estimated date-effects classified by stratum and dates.
2. Fit a normal regression model with date as the only term. The shape of date is modelled non-parametrically by a smoother.
3. The residual variance from this model is the stratum by date interaction; i.e., the herd size $\times$ date of slaughter interaction.

This estimate of the interaction is approximately invariant under different choices of dates where the effect is evaluated. This can be seen heuristically as follows.

The variance of the interaction effect is estimated as the residual variance from a regression of the interaction on date. If the number of date points where the effect is evaluated is large, then the addition of an extra point does not change the estimate of the stratum effect because the extra point will approximately be a linear interpolate from the two surrounding points. If the date effect initially is calculated in e.g. D points, the

addition of an extra point will on average approximately increase the sum of squares by a factor $(D+1)/D$ and the degrees of freedom from $(D-1)$ to $D(S-1)$, i.e. by a factor $D/(D-1) \approx (D+1)/D$. Thus little change will occur, if D is reasonably large, and if the points where the date effect is estimated are reasonably evenly spaced over the interval in which the original data lie.

As empirical evidence for this, we tried to estimate the date × herd size variation, using different points of evaluation of the date-effect in each of the 150 strata. We calculated the date effect at intervals of 3, 5 and 10 days (i.e. in 87, 53 and 27 points over a span of 130 days on either side of 1 July 1995). The resulting estimated residual standard deviations found were:

	Cutpoint	
Interval	OD 10	OD 30
3 days	0.0537	0.0770
5 days	0.0668	0.0957
10 days	0.0801	0.1203

There is an increasing tendency in the estimated standard error by increasing spacing of evaluation points. This is partly attributable to the fitting algorithm, a smoother (lo from S-PLUS function gam) that takes a fixed number of points rather than a fixed absolute span on the scale into the window used for smoothing. Thus, the fit tends to be better for the larger number of points where the date effect is evaluated.

References

Baggesen, D.L., Wegener, H.C., Bager, F., Stege, H., Christensen, J., 1996. *Salmonella enterica* infections in Danish slaughter pigs. Prev. Vet. Med. 26, 201–213.

Danske Slagterier, 1996. Statistik 1995. Axelborg, Axeltorv 3, 1609 Copenhagen V, Denmark.

Dewey, C.E., Wittum, T.E., Hurd, H.S., Dargatz, D.A., Hill, G.W., 1995. Herd and litter-level factors associated with the incidence of diarrhea morbidity and mortality in piglets 4–14 days of age. Swine Health Prod. 3, 105–112.

Kristiansen, K., Husum, P., Willeberg, P., 1985. Epidemiologiske analyser af nogle besatningsfaktorers betydning for forekomsten af kvagsalmonellose i Ringkøbing amt 1980–1983. Dansk Vet. 68, 695–703.

Mousing, J., Jensen, P.T., Halgaard, C., Bager, F., Feld, N., Nielsen, B., Nielsen, J.P., Bech-Nielsen, S., 1997. Nationwide *Salmonella enterica* surveillance and control in Danish slaughter swine herds. Prev. Vet. Med. 29, 247–261.

Nielsen, B., Baggesen, D.L., Bager, F., Haugegaard, J., Lind, P., 1995. 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. Vet. Microb. 47, 205–218.

Pritchard, G.C., 1987. Transmissible gastroenteritis in endemically infected breeding herds of pigs in East Anglia, 1981–85. Vet. Rec. 7, 226–230.

Svensmark, B., Nielsen, K., Willeberg, P., Jorsal, S.E., 1989. Epidemiological studies of piglet diarrhoea in intensively managed Danish sow herds. Acta Vet. Scand. 30, 55–62.

Wolfinger, R., O'Connell, D., 1993. Generalized linear mixed models: A pseudolikelihood approach. J. Stat. Comp. Sim. 48, 233–243.