Salmonella serosurveillance: Different statistical methods to categorize pig herds based on serological data

J. Cortiñas Abrahantes [a,*], K. Bollaerts [a], M. Aerts [a], V. Ogunsanya [a] and Y. Van der Stede [b]

[a] Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, Agoralaan 1, B–3590 Diepenbeek, Belgium

[b] Veterinary and Agrochemical Research Center (CODA/CERVA), Brussels, Belgium

Corresponding author:

José Cortiñas Abrahantes. Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, Agoralaan 1, B–3590 Diepenbeek, Belgium.

Tel.: +32 11268286; Fax: +32 11268299

E-mail: jose.cortinas@uhasselt.be

Abstract

This study proposes 3 different statistical methods that can be applied in order to categorize pig herds into two groups (high sero-reactors versus low sero-reactors) based on serological test results for *Salmonella*-specific antibodies in pigs. All proposed statistical methods were restricted to allocate about 10 % of the herds into the group defined by each of the statistical approaches as high level of sero-reactors. Previously, semi–parametric quantile regression has been used for this purpose, and here we compare it with two other alternatives, a naive method (based on the mean values) and a third one based on Activity Region Finder methodology in combination with random forest regression models. The serological response values (= the sample to positive ratio (S/P-ratio)) of 13,649 pigs from 314 Belgian pigs herds were used for this comparison. Around 14% of these herds were assigned to the high sero-reactor herd group by at least one of these three methods. The corrected level of agreement was calculated together with the pairwise agreement between all three methods in order to classify herds as high level or low level sero-reactors, resulting in an agreement level larger than 92 %. The results obtained from a fourth method, which is adopted by the Belgian Federal Agency for the Safety of the Food Chain (FASFC), was also compared to the previous three methods. The methods were compared in terms of their agreement as well as their advantages and disadvantages. Recommendations for each applied method are given in relation with the objectives and the required policy used to classify pig herds based on serological data.

Keywords: *Salmonella*, pigs, risk herds, semi-parametric quantile regression, activity region finder, random forest regression.

1. INTRODUCTION

Worldwide millions of *Salmonella* infections in humans have been reported every year, causing numerous deaths around the world (INFOSAN, 2005). In Belgium, Campylobacteriosis and Salmonellosis are the most frequently reported food–borne illnesses with 3693 reported cases of human salmonellosis in 2006 (Anonymous, 2006).

Public health concerns, political pressures and consumer demands have increasingly made prevention of *Salmonella* (Typhimurium) in pigs a priority. However, the control of *Salmonella* is very complex given the numerous potential sources of contamination. Thus, in order to establish an effective control of *Salmonella* it is essential to include and consider every step of the pork production chain. Therefore, primary production units have received special attention in control and surveillance program within the European Union (EU). In the EU regulation Nr 2003/99/EC and 2160/2003, the European Commission has set deadlines for its Member States to implement *Salmonella* surveillance programs in different livestock species that contribute to increase the risk of food borne infections in humans (Anonymous 2003a,b). To fulfill the surveillance obligations for pigs, most EU countries have already applied serological tests as a screening tool (Alban et al., 2002; Davies et al., 2003a,b; Berends et al., 1997; Osterkorn et al., 2001). However, serological surveillance for *Salmonella* has its limitations, given that the presence of antibodies is merely an indirect indicator of human health risk and it has been demonstrated its poor correlation with the presence of bacteria at the individual

63 animal level. Nevertheless, serological testing at herd level has been previously
64 applied to categorize pig herds into different 'risk' levels and subsequently to allocate
65 control measures for herds where bacteria are likely to be present (Alban et. al.,
66 2002a and Van der Gaag, 2004). Note that, the way pig herds are allocated into
67 different groups (levels) might differ between Member States. As an example the
68 British Zoonoses Action Plan (ZAP) classified pig herds as herds with high
69 seroprevalence (= ZAP2 and ZAP3 level) when more than 65% of at least 15 samples
70 were found positive according to a cut off level of 0.25 S/P-ratio (= Sample to
71 Positive ratio) using the Guildhay VETSIGN Kit (Davies et al., 2003). In Denmark, a
72 pig herd was categorized as 'risk' herd (level 2 and 3) when the 'serological
73 *Salmonella* index' (= a weighted average of the % of positive samples during 3
74 consecutive samplings) was higher than 40 (Alban et al., 2002b). A mix-ELISA test
75 (Nielsen et al., 1995) was used with a cut off level of 40% OD. These examples
76 illustrate the use of serological data, and the fact that several approaches can be used
77 to classify pig herds as possible risk herds for *Salmonella*. The Belgian Federal
78 Agency for the Safety of the Food Chain (FASFC) implemented in January 2005 a
79 National *Salmonella* serosurveillance and control program in pigs, which became
80 compulsory by means of a Royal and Ministerial act in July 2007 (Van der Stede et
81 al., 2007). The program aimed to categorize maximum 10% of the pig herds as risk
82 herds and they are obligated to implement follow up and control measures to reduce
83 *Salmonella* levels. The criteria applied to assign a pig herd as 'high risk' was: a pig
84 herd with mean S/P-ratio (mean value calculated from 12 samples per sampling)

equal or higher than 0.6 during 3 consecutive samplings. These 'high risk' herds are considered further in the paper as high sero-reactors (the high sero-reactor level group) herds.

The objective of this paper is to categorize pig herds into two groups (high or low sero-reactor) based on serological test results using several approaches: semi–parametric quantile regression, naive method, Activity Region Finder methodology and the implemented method by FASFC. In addition, the agreement between all these methods was assessed using two measures of concordance. A maximum of 10 % of the pig herds were allocated into the high sero-reactor level/group resulting from the approach used.

2. MATERIALS AND METHODS

2.1. THE SEROLOGICAL DATASET

Since 1 January 2005 blood samples are collected from growing and fattening pigs within the Belgian Aujeszky disease monitoring program (12 blood samples are collected from each pig herd in Belgium every 4 months) and they are also used to test for *Salmonella*-specific antibodies with an indirect ELISA (HerdChek Swine *Salmonella* Antibody Test Kit, Idexx Laboratories^{R}).

Ample detail about the data can also be found in Bollaerts et al., 2007. The dataset contained information from 314 pig herds, with a total of 13649 observations. The average number of pigs sampled per herd was 43. From those 314 herds less than 5 % (15) were sampled only once, while 90 % (277) of the herds were sampled at least 3

times. Around 68% of the herds (214) were sampled more than 4 times and 17% (53) more than 5 times.

The presence of *Salmonella*-specific antibodies in each sample was determined by indirect ELISA, relating the optical density (OD) values to the mean positive kit control through the sample to positive ratio (=S/P ratio) corrected with negative background values of the kit (S/P-ratio = OD_{sample} - $OD_{Neg\ Kit\ control}$ / $OD_{Pos\ Kit\ control}$ - $OD_{Neg\text{-}kit\ control}$). S/P-ratio values range in general between 0 and 4. Beside S/P-ratios other covariates such as herd_id number, sample size, sampling date, animal weight and sampling round were recorded as well. Animal weight was considered as a polychotomous categorical variable based on the weight category of the pigs (<40kg, 40–59kg, 60–79kg and >80kg) and the variable 'sampling time' was defined as a continuous variable calculated as the number of days between the sampling date and the starting date of the national *Salmonella* monitoring program in Belgium (=01/01/2005)).

2.2. STATISTICAL METHODS

In this section we will briefly describe some statistical methods that can be used to categorize herds according to the level of sero-reactors. More precisely, we will focus on three new procedures and the procedure already adopted by FASFC. The R statistical computing environment (Ihaka and Gentleman, 1996) and the R packages randomForest version 4.5–15 (Liaw and Wiener, 2002) and ARF version 1.0 (Amaratunga and Cabrera, 2004) is used.

2.2.1. Naive Approach (Method I)

The 'naive' approach, consist in calculating the herd-specific mean S/P ratio for the period in which the herd is followed-up. Subsequently, 10 % of the herds (= 32 herds of the total 314) with highest mean S/P-ratios are categorized as high sero-reactor herds.

2.2.2. Semi-parametric Quantile Regression Approach (Method II)

The second approach has been proposed by Bollaerts et al. (2007), it takes into account the sampling time and the weight of the animal to calculate a quantile curve which is used in a second stage to classify the animals above this quantile curve as risk animals. Later the herds are categorized in the high sero-reactor group based on a beta–binomial approach, which is used to calculate the probability that sero-reaction is higher in a particular herd compared to the other herds.

In this approach the high sero-reactor herds are identified based on the number of animals having (very) high S/P-ratios. This can be naturally modeled using semi-parametric quantile regression (QR). QR has the advantage over ordinary least squares regression of being more robust to outlying observations and as such, is more appropriate to analyze extremely skew distributed data (like S/P-ratios). Finally, high sero-reactor herds are categorized based on the number of ‘risk’ animals in the herd. To correct for the intra–herd correlation, corresponding beta–binomial p–values are calculated. More details can be found in Bollaerts et al. (2007).

2.2.2.1. Quantile curves of S/P-ratios

Upper quantile curves of animal S/P-ratios are estimated while accounting for confounding seasonal and animal age effects. In particular, the following semi–parametric model is used for each animal *i* measured at time *j* in herd *k*

$$\hat{SP}_{\theta_{ijk}} = h(sampling\ time)_{ijk} + I(weight)_{ijk} \quad (1)$$

with *h(.)* being a smooth P–splines function (Eilers and Marx, 1996), *I* being an indicator matrix and $\hat{SP}_{\theta_{ijk}}$ the estimated quantile curve, where θ is indicating the quantile that we are interested in. Risk animals are defined as animals for which the observed S/P-ratio is higher than the corresponding θ×100% quantile. For this particular case θ was chosen to be 0.9.

2.2.2.2. Beta–binomial p–values

In order to categorize the herds in the high sero-reactor group, the number of risk animals Y_k is calculated for each herd *k*. Then, using the beta–binomial distribution for correlated binary data, the number of risk animals Y_k in herd *k* is beta–binomially distributed ($Y_k \sim BB(n_k, 1-\theta, \rho)$), with n_k being the total number of sampled pigs in herd *k*, $1-\theta$ the probability of being a risk animal and ρ being the intra–herd correlation. The p-value can then be calculated as $p_k = P\{Y_k \geq y_k \mid Y_k \sim BB(n_k, 1-\theta, \rho)\}$, low p-values indicating herds belonging to the high sero-reactor group.

2.2.3. Random Forest (RF) in combination with Activity Region Finder (ARF) methodology (Method III)

The third approach uses the residuals obtained from the Random Forest regression models (Breiman, 2001) in combination with a tree based method (Activity Region Finder, mainly employed in microarray experiments to identify highly expressed genes, Amaratunga and Cabrera, 2004). In the first stage residuals from the random forest regression model of the S/P ratio against the sampling date and the weight category of the pigs are obtained. In a second stage, these residuals are then used to define high activity regions via ARF methodology. A total of 1000 trees were grown in the random forest analysis, extra information about the importance of the covariates used in the model to explain the S/P ratio behavior is obtained by two importance measures: the first one based on the reduction of mean squared error (MSE) and the second one based on the total decrease in node impurities from splitting on the variable, averaged over all 1000 trees. These importance measures define how strong the relation is between the covariate and the response, and which covariates are stronger associated with the S/P ratios.

2.2.3.1. Random Forest

The random forest method (Breiman, 2001) is a supervised learning algorithm; it is an ensemble of many identically distributed trees generated from bootstrap samples of the original data. Each tree is constructed via a tree classification algorithm. The simplest random forest with random features is formed by selecting randomly, at each node, a small group of input variables to split on. The size of the group is fixed

191 throughout the process of growing the forest. Each tree is grown by using CART
192 methodology without pruning (Breiman et al., 1984). Also, two measures of variable
193 importance can be obtained, the first one associated to the reduction in mean squared
194 error (MSE) using the out of bag portion (the part of the data left out in each
195 bootstrap sample) and another one which quantifies the decrease in node impurity (is
196 the total decrease in node impurities (measured using Gini index or residual sum of
197 squared) from splitting on the variable, averaged over all trees).

198 *2.2.3.2. Activity Region Finder Methodology (ARF)*

199 This methodology is similar to conventional classification tree methods. ARF uses a
200 so called *H*–criterion in order to identify the High Activity Regions (HARs). This *H*–
201 criterion focuses entirely on the incidence rate of successes within an interval, in such
202 a way that the subset selected at each step will have a substantial (and rapidly
203 growing) proportion of successes. Accordingly, intervals at step (*s*) would have
204 substantially higher proportion of success than intervals in the previous step ($s - 1$).
205 Two methods of recursive partitioning based on H–criterion (ARF1 and ARF2) have
206 been proposed by Amaratunga and Cabrera (2004).

207 ***2.2.4. FASFC Adopted Method (Method IV)***

208 The last approach, which is adopted by FASFC, consists in monitoring the herd-
209 specific mean S/P ration per sampling round. A herd is assigned to the high sero-
210 reactor group if during 3 consecutive sampling rounds the herd-specific mean S/P
211 ratio is above 0.6 (Van der Stede et al., 2007).

212 ***2.2.5. Evaluation of Agreement Between the four Methods***

213 The agreement between each two of methods, and for each category, together
214 with overall agreement is estimated. The overall level of agreement (**kappa (κ)**
215 statistic) for all methods (Naive **κ**, which compares the agreements to that expected if
216 the methods were independent (Cohen, 1960)) is calculated. To illustrate the
217 procedure used to calculate the level of agreement, we consider a 2×2 contingency
218 table, for two of the methods proposed. Then $\kappa = \frac{p - e(\kappa)}{1 - e(\kappa)}$, where p denotes the
219 overall agreement propensity (proportion of the 314 herds for which high and low
220 sero-reactor group categorization coincide for both methods) and $e(\kappa)$ the
221 propensity of both raters (methods) to agree by chance without having the same
222 assessment of a herd. The correction for chance agreement is then calculated
223 as $e(\kappa) = \frac{A_1}{N} \cdot \frac{B_1}{N} + \frac{A_2}{N} \cdot \frac{B_2}{N}$, where A_1, A_2, B_1 and B_2 are the marginal totals in the 2×2
224 table, *A* or *B* indicate the methods, the sub-indices indicate the low or high sero-
225 reactor group and *N* indicates the total number of herds (314).
226 If a rating is random, it can be demonstrated that agreement can occur with a
227 fixed probability of 0.5. It follows then that chance agreement probability should not
228 exceed this value. Gwet (2002), shows that $e(\kappa)$ can exceed 0.5 and proposed a
229 correction to the agreement measure. In this paper, we also present the corrected
230 measure (Corrected **κ***: $\kappa^* = \frac{p - e(\kappa^*)}{1 - e(\kappa^*)}$) based on the procedure proposed by Gwet
231 (2002)). This measure appropriately corrects the agreement propensity for chance

agreement using the following expression $e(\kappa^*) = \frac{A_1 + B_1}{N} \cdot \left(1 - \frac{A_1 + B_1}{2 \cdot N}\right)$. It was also demonstrated (Gwet, 2002) that the corrected measure of chance agreement probability never takes values above 0.5.

3. RESULTS

For Method I, the herds with herd–specific mean S/P ratios larger than 0.898 (given that we should categorized maximum 10 % of the herds (at most 32 herds out of 314) as high sero-reactors) are categorized to belong to the high sero-reactor group. If instead, Method II was applied, then those 32 herds having the smallest beta-binomial probabilities (beta-binomial probability smaller than 0.00056) are categorized into the high-sero-reactor group. The third approach (Method III: ARF combined with random forest), provide also information about variable importance and the fit of the model. The multiple determination coefficient obtained from the random forest analysis was 0.426. The resulting tree obtained from the ARF procedure is shown in **Figure 1**. In each of the significant nodes detected by the ARF procedure we can see the number of observation in the node (*n*) and the mean of the residuals for the observations in the node (*M*). Note also, that in each node there is a reference to a category, which is just a set of herd identification numbers present in each node. From this figure it is clear that the herds classified in the first split, the right branch (with 2007 observations from 50 herds) have higher residuals compared to the rest of the herds. This means that even when the effects of the covariates are removed, still some systematic effect is present in our data. Therefore, other factors

definitely influence the observed values, indicating that these factors should be followed up. In order to be able to compare the results from this method (Method III) with the other two approaches we select from the 50 herds with high residuals, those 32 with the highest residual mean (residual mean above 0.224), which are then considered to be high level sero-reactors herds. The FASFC method (Method IV) (**Table I**) identified 28 risk herds (8.9%) as high sero-reactors herds. From those, 20 were also categorized as high sero-reactor herds by one of the 3 methods here presented (highlighted **in bold** in **Table I**). In total 52 herds (of the 314 herds) were categorized as high sero-reactors herds by at least one of the 4 different methods, about 21 % (11 herds) were classified as high sero-reactors by the four methods. From the other herds (41), 9 were classified by one of the three new approaches, 6 of them were selected by Method I, Method IV and one of the other two approaches, 5 by Method II, Method IV and one of the other two approaches and 3 selected by Method III, Method IV and one of the other two approaches.

As it was mentioned before a total of 52 herds were categorized as high sero-reactors herds by at least one of the four methods (**Table I**), the remaining 262 herds were classified as low sero-reactor herds by all four methods. Around 67 % of the herds are classified as high by at least two of the four methods. If we focus on those herds that are categorized as high for at least 3 of the four methods, then 50 % of the herds (26) satisfied this condition.

Table II, shows the variable importance measures obtained from the random forest analysis, both measures of variable importance produce larger value for the

variable sampling time. In terms of means squared error (MSE) reduction, both variables reduced the MSE in more than 70%, producing comparable results, but in terms of decrease of node impurity, sampling time produces a considerably larger decrease. This indicates a strong seasonal character of *Salmonella* infections in pig herds.

The agreement between the applied methods is presented in **Table III**. The overall level of agreement for the 3 new methods without correction is around 97 %, indicating that around 97 % of the classifications of herds are in agreement for all three approaches. When the correction factor is taken into account the level of agreement decreases towards 94 % but is still very high. The level of agreement is also high considering the herds with low level of sero-reactors. However, the corrected level of agreement for all three new methods, when we focus on the high level group, is around 71 %. When we compare all four methods, still high overall corrected level of agreement (around 92 %) is found, but again, if we focus on the high level group, the percentage of agreement drops to 56%.

Figure 2 shows the percentage of the 44 selected herds that are classified by one, two or all three methods applied (FASFC method is not considered). It can be seen that 21 herds (48%) are selected by all three methods as high sero-reactors herds while 4 (9%), 4 (9%) and 2 herds (5%) are selected by respectively Method I and Method II, Method I and Method III and Method II and Method III. Thirteen herds (30%) were selected by only one method: 3 (6.82%), 5 (11.36%) and 5 herds (11.36) for respectively Method I, Method II and Method III.

297 A cut-off value of 15% is also used to divide the population of herds into two
298 groups. It can be seen that a similar level of agreement was found (**Table IV**). The
299 overall level of corrected agreement is close to 90%.

300 **4. DISCUSSION AND CONCLUDING REMARKS**

301 **4.1. Example 1: Herd categorized only by Method I**

302 The first approach (Method I) takes into account neither the sampling time, nor
303 the weight of the animal even though they are known confounding factors. S/P-ratios
304 are expected to be higher during the summer and autumn and are also expected to be
305 higher for fattening pigs compared to weaners (<40kg) and growing pigs between 40
306 and 60 kg (Van der Stede et al., 2007).

307 For example, the herd with ID '46' is only selected using Method I (**Table I**),
308 from the 36 pigs sampled on this herd (all from weight category >80kg at 3 different
309 sampling times 310, 453 and 582) only 9 (1,02, 1.52, 2.02, 2.21, 2.90, 3.08, 3.16,
310 3.17 and 3.22) of them are above the specified quantile used to classify animals at
311 risk, 19 animals had a S/P ratio below 0.7 and 8 animals had an S/P-ratio between 0.7
312 and 0.9. This example shows that the mean S/P-ratio at herd level might be inflated
313 due to a low number of animals having a (very) high individual S/P-ratio, which may
314 be linked with the fact that the weight of the animals sampled were above 80 kg (old
315 animals). However, high S/P-ratios certainly indicate that a *Salmonella* infection had
316 occurred in that herd. Moreover, a high association between bacteriological isolation
317 of *Salmonella* spp. in faeces and the mean S/P-ratio at herd level has been already
318 reported (Laevens et al. (2005) unpublished data). The other three approaches did not

select this herd. However, the probability obtained with Method II is also very small and the mean residual from the random forest (ARF) approach is close to the threshold value used. The variances of the residuals for the herds that were only selected by Method I (around 0.7) were in general higher than those herds that were classified only by either Method II or Method III.

Another important distinction between Method I, Method II and III, is that the two last do take into account the sampling time and the animal weight, but they proceed in different ways. The semi-parametric quantile regression approach focuses on the animal level and classifies each animal at risk or not. Subsequently, this method computes the proportions of risk animals in a herd and finally, using a beta–binomial approach to correctly account for correlation between animals in a herd and selects the herds with higher antibody levels compared to the other herds. The Random Forest (ARF) approach, basically transforms the response, in a sense that the effect of sampling time and animal weight on the S/P ratios is removed, and uses it to select those herds which have higher values of the residuals, meaning that systematic patterns are still present in the data, and not yet removed.

4.2. Example 2: Herd categorized only by Method II

The herd with ID ‘40’ (**Table I**) is only selected by Method II. This herd was sampled twice with a total of 24 pigs with weights between 60–79kg. From the 24 animals, 9 (5 of them being growing pigs) had S/P ratio values above the quantile curve (S/P-ratio above 1), 10 animals had S/P ratio below 0.6 and 5 animals with a S/P-ratio between 0.6 and 0.95. It is clear that the proportion of animals classified as

risk animals by Method II is higher for this herd (0.375) than for the herd with ID '46'. If we focus on Method III more negative residuals with relative moderate variance (in general around 0.5) is observed. The mean S/P-ratio for this herd was around 0.83 and below the threshold of 0.898.

4.3. Example 3: Herd categorized by Method III

On the other hand, herd with ID '41' was selected as risk herd only by Method III. In that herd 48 pigs in total were sampled during 4 consecutive sampling rounds, and weights varied between 60 and 110 kg. Only 5 animals at that herd had S/P ratios above the quantile curve, but 38 % of the animals had S/P-ratios higher than 0.9, half of them from weight category 60–79kg. Twenty-two percent (11/48) had S/P-ratios between 0.2 and 0.5 and there were no animals with S/P-ratios below 0.2. For this herd we observed that around 80% of the residuals are having positive sign with small variability (in general around 0.2). The mean S/P-ratio was 0.80, relatively close to the threshold value (0.898).

4.4. Example 4: Herd selected only by Method IV

Using the method applied now by the FASFC (Method IV), 8.9 % (28/314 herds) of the herds were assigned to the high sero-reactor group from which 20 (71%) were also selected by one of the 3 other methods. In order to illustrate the differences with respect to the other three methods discussed in this paper, we will focus our attention to one of the herds classified only by the approach adopted by FASFC (Herd ID '52'). In this herd 60 animals in total were sampled during 5 consecutive sampling rounds: in round 1 (time point 62) animals with weights below 60kg and above

80kg were sampled, in round 2 and 4 (time point 190 and 449) pigs with weights below 60kg and pigs with weights between 60 and 80kg were sampled, in round 3 and 5 (time point 323 and 573) only weaners (pigs with weights below 40 kg) were sampled. The mean S/P ratio profile of this herd was 0.272, 0.995, 0.625, 0.605 and 0.502. The herd is assigned to the high sero-reactor group by Method IV as it showed a S/P ratio above 0.6 during sampling rounds 2, 3 and 4. However, in sampling round 4 only fattening pigs were sampled, and it is known that their S/P ratios are in general high. In fact, at that moment only 1 out 12 animals had S/P ratio larger than 1, 6 animals with S/P-ratio between 0.24 and 0.5 and 5 pigs having S/P-ratios between 0.58 and 0.75. From the 60 animals sampled, 63% had S/P-ratios below 0.6. Only 7 (12%) animals were classified as 'risk' animals by Method II, for a beta-binomial probability of 0.393, much larger than the specified one (0.00056). The overall mean S/P-ratio is 0.599 which was substantially lower than the threshold value (0.898) proposed by Method I. Finally, the mean of the residuals (0.065) obtained in Method III is also smaller than the 0.224 proposed. This indicates that Method IV does not account for confounding factors and it merely focus on the average S/P ratio of consecutive samplings. It is well known how the mean can be influenced by a small number of observations with very high values. It is also important to highlight that Method IV will not categorize herds in the high sero-reactor group, if oscillating behavior is observed (jumping above and under the cut-off level of 0.6).

4.4. Level of Agreement between the Methods

A high overall level of agreement between the four methods was observed (above 91 %). Thirty-five herds are selected by at least two of the methods. However, 17 other herds have been selected by only one of the approaches, indicating that the selection of herds to be followed up should be done very careful. The choice for the most appropriate method depends clearly on the purpose and the policy of interest. If interest is in risk animals having very high S/P ratios, then the semi-parametric quantile regression approach (Method II) is preferred. If instead, we focus on herd-level, as is done in many *Salmonella* surveillance programs, the random forest (ARF) approach (Method III) and/or the naive procedure (Method I) can be used. The advantages, applications, as well as the limitations of each of these methods are summarized in **Table V**.

The high level of agreement is of practical importance as it implies that categorization of herds with high antibody levels is a consistent tool. We agree upon the fact that firm proof of being a herd with high level of sero-reactors can only be assessed by a positive culture of a faeces sample and/or lymph node (at slaughterhouse level). However, due to non perfect sensitivity of the bacteriological isolation method for *Salmonella* (ISO 6579 (Annex D) method) and its sampling procedure at herd level, the problem of correlation between serology and bacteriology will always exist, no matter which 'serological' method is used to categorize herds.

In conclusion, the categorization of herds with high level of sero-reactors is consistent and a high level of agreement between the methods was observed.

Another important issue is the choice of the cut-off value to separate the population of herds into a high and a low group of sero-reactors herds. We have considered two different values (10 and 15%) and have shown that there is little impact on the overall κ^*. Of course, to have a rigorous evaluation of the impact further research is needed, but these results already point towards a rather limited impact on the choice of the cut-off value with respect to the level of agreement between the methods.

The use of other methods that can correctly account for confounding factors such as pig weight and seasonal effects and possible cluster effects, and/or adapting sampling plans is a subject of further research.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the fund of Scientific Research (FWO, Research Grant # G.0151.05) and Belgian IUAP/PAI network # P6/03 ``Statistical Techniques and Modeling for Complex Substantive Questions with Complex Data" of the Belgian Government (Belgian Science Policy).

REFERENCES

Alban, L., Olsen, A.M., Nielsen, B., Sorensen, R., Jessen, B., 2002a. Qualitative and quantitative risk assessment for human salmonellosis due to multi-resistant *Salmonella Typhimurium* DT104 from consumption of Danish dry-cured pork sausages. Prev. Vet. Med. 52, 251–265.

Alban, L., Stege, H., Dahl, J. 2002b. The new classification system for slaughter-pig herds in the Danish *Salmonella* surveillance-and-control program. Preventive Veterinary Medicine 53, 133 – 146.

Anonymous 2003a. Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the control of *Salmonella* and other specified food-borne zoonotic agents.

Anonymous 2003a. EC-Zoonoses Directive (2003/99/EC) of the European Parliament of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC.

Anonymous 2003b. EC-Zoonoses Directive (2003/99/EC) of the European Parliament of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC.

444 Anonymous 2006. Trends and Sources. Report on zoonotic agents in Belgium 2006.
445 Working group on foodborne infections and intoxications. Federal Agency for
446 the Safety of the Food Chain (FASFC), Scientific Institute of Public Health
447 (WIV- IS/P), Veterinary and Agrochemical Research Centre (CODA-CERVA).

448 Berends, B.R., Van Knapen, F., Snijders, J.M., Mossel, D.A.1997. Identification and
449 quantification of risk factors regarding *Salmonella* spp. on pork carcasses. Int J
450 Food Microbiol 36, 199 – 206.

451 Amaratunga D., and Cabrera J., 2004. Mining Data to Find Subsets of High Activity.
452 Journal of Statistical Planning and Inference 122, 23 – 41.

453 Bollaerts K., Aerts M., Ribbens S., Van der Stede Y., Boone I., and Mintiens K.,
454 2007. Identification of *Salmonella* Risk Herds in Belgium using semi–parametric
455 quantile regression. Journal of the Royal Statistical Society: Series A 171, 449 –
456 464

457 Breiman L., 2001. Random forests, Machine Learning 45, 5 – 32.

458 Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984. *Classification and*
459 *Regression Trees*. Chapman & Hall, New York.

460 Cohen J., 1960. A coefficient of agreement for nominal scales. Educational and
461 Psychological Measurement 20, 37 – 46.

462

463 Collard J.M., Bertrand S., Willems L., Baeyens D., De Cooman F., Steenhaut H.,
464 Lattuca M., Mairiaux E., Dupont Y., Godard C., Wildemauwe C. and Vrints M.,
465 2003. Human salmonellosis in Belgium: Trends in 2003. National Reference
466 Center for *Salmonella* and Shigella, Belgium.
467 Collard JM, Bertrand S, Dierick K, Godard C, Wildemauwe C, Vermeersch K,
468 Duculot J, Van Immerseel F, Pasmans F, Imberechts H, Quinet C. 2008. Drastic
469 decrease of *Salmonella* Enteritidis isolated from humans in Belgium in 2005,
470 shift in phage types and influence on foodborne outbreaks. Epidemiol Infect.136,
471 771 – 81.

472 Davies, R.H., Cook, R. 2003a. The UK ZAP *Salmonella* Scheme for pig meat
473 production (Weybridge).
474 Davies R.H., Heath P.J., Coxon S.M., Sayers A.R., 2003b. Evaluation of the use of
475 pooled serum, pooled muscle tissue fluid (meat juice) and pooled faeces for
476 monitoring pig herds for *Salmonella*, J. Appl. Microbiol 95, 1016 – 1025.
477 Eilers, P.H.C., and Marx, B.D., 1996. Flexible smoothing using B–splines and
478 penalized likelihood (with comments and rejoinder). Statistical Science 11 (2),
479 89 – 121.
480 Gwet K., 2002. Kappa statistic is not satisfactory for assessing the extent of
481 agreement between raters. Statistical Methods for Inter–Rater Reliability
482 Assessment Series 1, 1 – 16.
483

Ihaka R. and Gentleman R., 1996. R: A Language for data analysis and graphics. Journal of Computional and Graphical Statistics 5, 299 – 314.

International Food Safety Authorities Network (INFOSAN), 2005. Antimicrobial resistant *Salmonella* [on line] http://www.doh.gov.za/department.

Institutional Animal Care and Use Committee, Emory University, USA [on line] (1996) http://www.research.ucsb.edu/connect/pro/disease.html.

Laevens, H., Mintiens, K., 2005, De associatie tussen *Salmonella* serologie en de isolaten van *Salmonella* bij slachtvarkens op Belgische varkensbedrijven. Report FASFC 2005, 1 – 25.

Liaw A. and Wiener M., 2002. Classification and regression by random forest. The Newsletter of the R Project 2/3, 18 – 22.

Nielsen, B., Baggesen, D., 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 Microbiol 47, 205 – 218.

Nollet, N., Maes, D., Duchateau, L., Hautekiet, V., Houf, K., Van Hoof, J., De Zutter, L., De Kruif, A., Geers, R., 2005. Discrepancies between the isolation of *Salmonella* from mesenteric lymph nodes and the results of serological screening in slaughter pigs. Vet Res 36, 545 – 555.

Osterkorn, K., Czerny, C.P., Wittkowski, G., Huber, M., 2001. Sampling plan for the establishment of a serologic *Salmonella* surveillance for slaughter pigs with meat juice ELISA. Berl Munch Tierarztl Wochenschr 114, 30 – 34.

Rautiainen, E., Konradsson, K., Lium, B., Mortensen, S., Wallgren, P., 2001 Disease Surveillance Strategies in Swine. Acta vet. scand. 2001, Suppl. 94, 31-42.

Van der Gaag, M.A., 2004. Epidemiological and economic simulation of *Salmonella* control in the pork supply chain. PhD Thesis. Wageningen University, Wageningen.

Van der Stede Y., Boone, I., Bollaerts, K., Herr, C., Cortiñas Abrahantes, J. Aerts, M., Hooyberghs, J. Vermeersch, K., Imberechts, H., Mintiens, K. 2007. Implementing a *Salmonella* Surveillance an control program in Belgian pigs herds. Proceedings from VEE Studiedagen 12-13 december.

Stege H., Christensen J., Nielsen J.P., Baggesen, D.L., Enoe C., Willeberg P., 2000. Prevalence of subclinical *Salmonella* enterica infection in Danish finishing pig herds. Prev. Vet. Med. 44, 175 – 188.

Wilm K.H., 2005. Our Food. Database of Food and Related Sciences, Germany [on line] http://www.ourfood.com.

FIGURE LEGENDS

Figure 1. Significants Nodes of ARF Tree for residuals obtained from the random forest model. In each of the significant nodes detected by the ARF procedure the number of observation in the node (*n*) and the mean of the residuals for the observations in the node (*M*) is displayed. In each node a reference to a category (a set of herd identification numbers) is given.

Figure 2. Percentage of herd classified for one or more procedures from the 62 identified by at least one of the methods.

FIGURES

Fig 1.

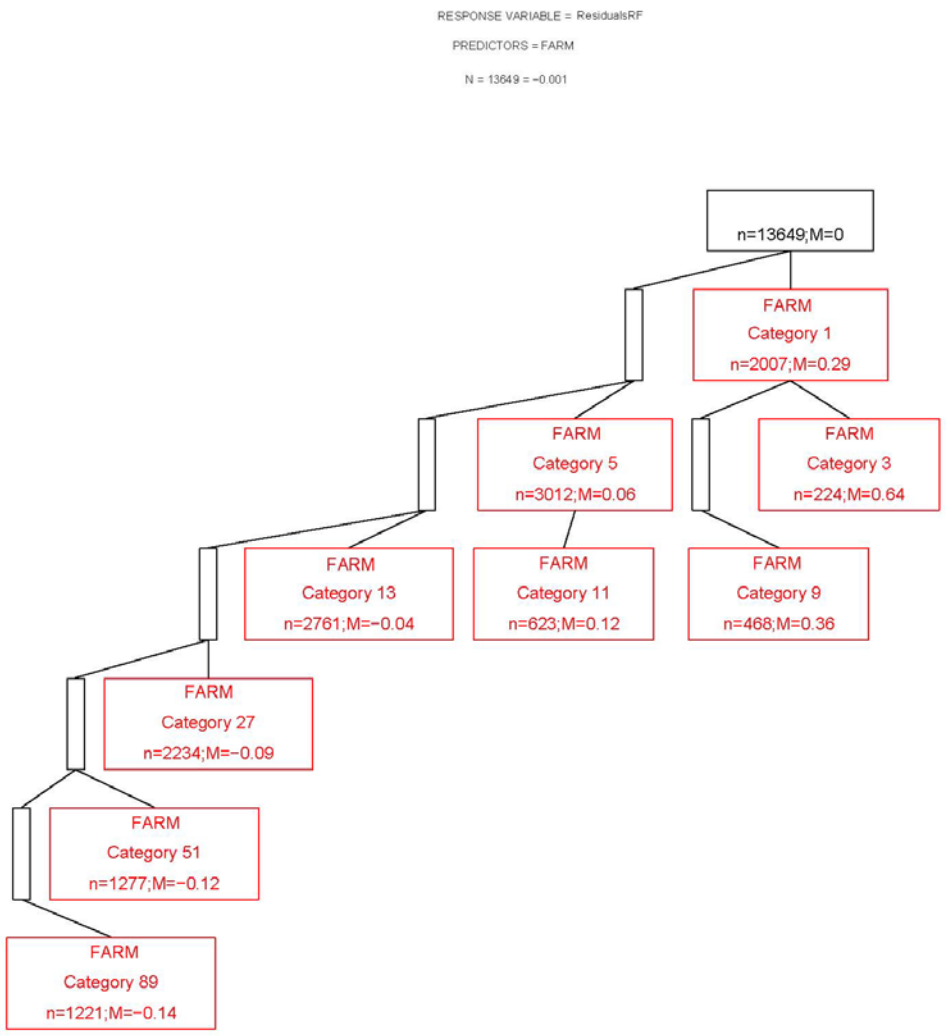
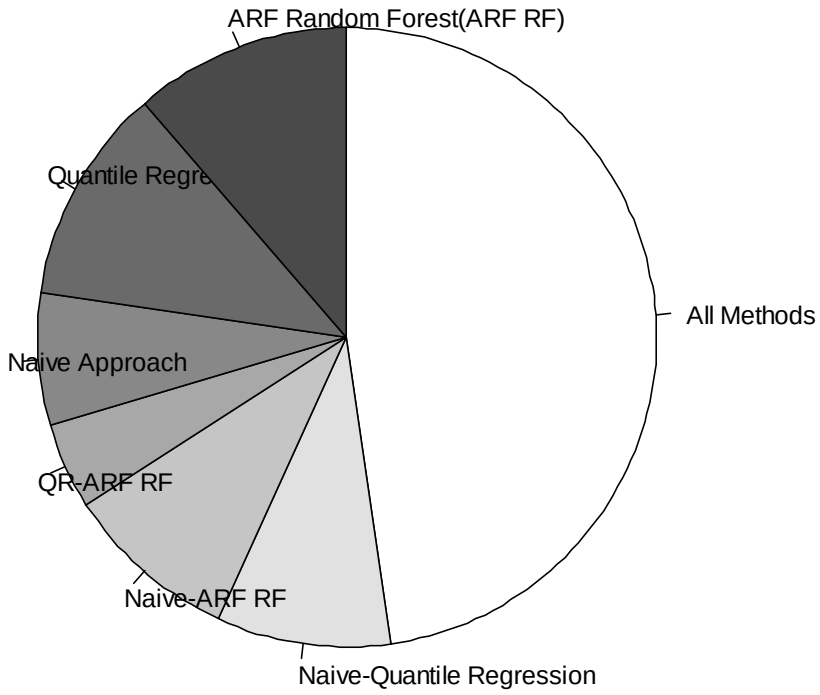


Fig 2.


ARF Random Forest(ARF RF)
Quantile Regre
All Methods
Naive Approach
QR-ARF RF
Naive-ARF RF
Naive-Quantile Regression

TABLES

Table 1

Herds classified by one of the four methods (M1: Naive Approach (mean SP–ratio), M2: Quantile Regression (beta–binomial probability), M3: Random Forest–ARF (mean of the residuals) M4: FASFC method (number of times sampled; number of times S/P ratio above 0.6)) as a herd with high level of sero-reactors (1) and low level of sero-reactors (0) and the number of times a herd was categorized as high level. The highlighted herd ID's are those herds that are also identified by the method adopted by FASFC as described in the introduction part.

Herd ID	M1	M2	M3	M4	Times	Herd ID	M1	M2	M3	M4	Times
1	**1(1.59)**	**1(8.9E-19)**	**1(0.62)**	**1(3;3)**	**4**	27	0(0.89)	1(2.5E-05)	1(0.33)	0(3;2)	2
2	**1(1.68)**	**1(6.7E-15)**	**1(0.73)**	**1(4;4)**	**4**	28	1(1.10)	0(6.1E-04)	1(0.59)	0(2;1)	2
3	**1(1.44)**	**1(9.3E-15)**	**1(0.62)**	**1(4;4)**	**4**	**29**	**0(0.88)**	**0(6.6E-04)**	**1(0.43)**	**1(4;3)**	**2**
4	**1(1.15)**	**1(2.5E-14)**	**1(0.26)**	**1(5;5)**	**4**	**30**	**1(0.92)**	**0(6.7E-04)**	**0(0.14)**	**1(5;3)**	**2**
5	**1(1.18)**	**1(3.1E-13)**	**1(0.37)**	**1(4;4)**	**4**	31	1(0.91)	0(2.7E-03)	1(0.22)	0(4;3)	2
6	**1(1.47)**	**1(2.9E-12)**	**1(0.35)**	**1(4;4)**	**4**	**32**	**1(0.92)**	**0(6.8E-03)**	**0(0.17)**	**1(4;3)**	**2**
7	**1(1.36)**	**1(1.3E-08)**	**1(0.48)**	**1(3;3)**	**4**	33	1(1.42)	0(2.6E-02)	1(0.23)	0(1;1)	2
8	**1(1.52)**	**1(1.0E-07)**	**1(0.63)**	**1(3;3)**	**4**	34	1(1.24)	0(4.7E-01)	1(0.71)	0(1;1)	2
9	**1(1.12)**	**1(2.5E-05)**	**1(0.23)**	**1(3;3)**	**4**	**35**	**0(0.83)**	**0(5.3E-01)**	**1(0.34)**	**1(4;4)**	**2**
10	**1(0.92)**	**1(4.4E-05)**	**1(0.25)**	**1(4;3)**	**4**	36	0(0.70)	1(5.8E-07)	0(0.13)	0(5;2)	1
11	**1(0.92)**	**1(4.4E-05)**	**1(0.31)**	**1(4;3)**	**4**	37	0(0.79)	1(5.6E-05)	0(0.14)	0(5;2)	1
12	1(1.90)	1(1.6E-15)	1(0.88)	0(2;2)	3	38	0(0.89)	1(1.2E-04)	0(0.11)	0(3;2)	1
13	**1(1.27)**	**1(4.4E-14)**	**0(0.15)**	**1(5;3)**	**3**	39	0(0.53)	1(2.4E-04)	0(0.08)	0(5;2)	1
14	1(1.67)	1(3.3E-11)	1(0.35)	0(2;2)	3	40	0(0.83)	1(3.2E-04)	0(0.15)	0(2;2)	1
15	1(1.23)	1(2.0E-10)	1(0.31)	0(4;2)	3	41	0(0.80)	0(6.1E-04)	1(0.24)	0(5;2)	1
16	1(1.24)	1(1.5E-09)	1(0.28)	0(4;2)	3	**42**	**0(0.69)**	**0(2.0E-03)**	**0(0.07)**	**1(5;3)**	**1**
17	**1(1.24)**	**1(3.7E-07)**	**0(0.21)**	**1(4;4)**	**3**	**43**	**0(0.72)**	**0(5.7E-03)**	**0(0.11)**	**1(5;3)**	**1**
18	1(1.42)	1(1.0E-06)	1(0.55)	0(1;1)	3	**44**	**0(0.62)**	**0(5.7E-03)**	**0(0.08)**	**1(5;3)**	**1**
19	**1(0.95)**	**1(2.0E-06)**	**0(0.21)**	**1(4;4)**	**3**	**45**	**0(0.85)**	**0(6.8E-03)**	**0(0.20)**	**1(4;4)**	**1**
20	1(1.07)	1(5.2E-06)	1(0.31)	0(4;3)	3	46	1(1.02)	0(7.7E-03)	0(0.21)	0(3;2)	1
21	1(1.10)	1(5.2E-05)	1(0.44)	0(2;2)	3	47	0(0.75)	0(1.9E-02)	1(0.25)	0(4;2)	1
22	1(1.14)	1(8.8E-05)	1(0.29)	0(4;2)	3	**48**	**0(0.71)**	**0(1.9E-02)**	**0(0.16)**	**1(4;3)**	**1**
23	1(0.90)	1(1.2E-04)	1(0.31)	0(3;2)	3	49	0(0.87)	0(2.8E-02)	1(0.39)	0(2;2)	1
24	**1(0.90)**	**1(1.8E-04)**	**0(0.21)**	**1(4;3)**	**3**	**50**	**0(0.77)**	**0(1.0E-01)**	**0(0.17)**	**1(4;3)**	**1**
25	1(1.23)	1(5.4E-04)	1(0.27)	0(1;1)	3	**51**	**0(0.65)**	**0(1.4E-01)**	**0(0.11)**	**1(5;3)**	**1**
26	**0(0.89)**	**1(5.6E-04)**	**1(0.25)**	**1(3;3)**	**3**	**52**	**0(0.60)**	**0(3.9E-01)**	**0(0.07)**	**1(5;3)**	**1**

Table 2

Importance measures based on MSE (%) and node impurity obtained from the random forest analysis as described in section *2.2.3.1*.

Variable	Reduction in MSE (%)	Decrease in Node Impurity
Sampling Time	75.78	289.34
Weight Category	73.16	150.35

Table 3

Pairwise and overall level of agreements when 10% of the herds are categorized as high sero-reactors.

Methods	Naive κ (%)			Corrected κ* (%)		
	Low	High	Overall	Low	High	Overall
M1 vs M2	97.52	78.12	95.54	96.96	73.22	94.54
M1 vs M3	97.52	78.12	95.54	96.96	73.22	94.54
M2 vs M3	96.81	71.88	94.27	96.1	65.57	92.98
M1, M2 and M3	97.28	76.04	95.12	96.67	70.67	94.02
M1 vs M4	94.76	60.71	91.72	93.67	52.51	89.99
M2 vs M4	94.41	57.14	91.08	93.24	48.19	89.22
M3 vs M4	93.71	50.00	89.81	92.4	39.55	87.68
ALL Four Methods	96.11	64.52	92.99	95.27	56.18	91.48

Table 4

Pairwise level of agreements when 15% of the herds are categorized as high sero-reactors.

Methods	Naive κ (%)			Corrected κ* (%)		
	Low	High	Overall	Low	High	Overall
M1 vs M2	96.62	81.25	94.27	95.44	74.7	92.26
M1 vs M3	97.74	87.50	96.18	96.95	83.13	94.84
M2 vs M3	95.11	72.92	91.72	93.4	63.45	88.83
M1 vs M4	90.21	71.43	88.54	87.56	63.71	85.44
M2 vs M4	90.56	75.00	89.17	88.01	68.24	86.25
M3 vs M4	90.56	75.00	89.17	88.01	68.24	86.25

Table 5

Summary of limitations, advantages and situations in which each of the methods could be applied.

	Naïve	Semi-Parametric Quantile Regression	ARF-Random Forest	FASFC Adopted method
When	Focus on mean behaviour (general herd level surveillance)	Focus on 'Salmonella infected' herds (herds with a history of consecutive high S/P ratios in all samplings)	Focus on those herds with high mean S/P ratios after correcting by confounding factors	Focus on those herds with consecutively high mean S/P ratios for at least three visits
Limitations	Do Not correct for confounding factors. Do not correct for any correlation and/or association between sampled animals at a particular time point. Sensitive for extreme values.	Do not correct for any correlation and/or association between sampled animals at a particular time point. The model that correct for possible confounding factors should be prespecified in advance. Complex to install and perform in practice (statistical tools).	Do not take into account possible associations. Complex to install and perform in practice (statistical tools).	Do not take into account possible associations. Do Not correct for confounding factors. Sensitive for extreme values. Discard herds with oscillating behavior (mean S/P ratio jumping above and bellow cut-off (0.6))
Advantages	Simple to implement and explain (herders, veterinarians). Easy follow up. Fast. No need for sophisticated statistical tools.	Flexible model for time effects. Possibility of using different upper quantiles for θ *(changing criteria is flexible once system is installed).* Less sensitive for extreme values (robust technique).	Do not need prespecification of type of relationship between S/P ratios and covariates.	Simple to explain and implement (herders, veterinarians). Easy follow up. Fast. No need for sophisticated statistical tools.