Statistical Aspects in the Styrian Salmonella surveillance program in pigs

Fuchs K¹, Koefer F², and Pless P²

¹Institute of Applied Statistics, Joanneum Research, Steyrergasse 25a, 8010 Graz, Austria
²Department of Veterinary Administration in Styria, Zimmerplatzgasse 15, 8010 Graz, Austria

Abstract

Surveillance and Monitoring programs to substantiate freedom of disease are becoming increasingly important. Sample sizes and the right selection of the samples play a decisive role. In the course of this paper the sampling schemes for the Styrian Salmonella surveillance program in pigs are presented. The influence of diagnostic tests on the result of the estimation of prevalence is discussed and different sample sizes depending on the sensitivity and the specificity of the used diagnostic test are calculated.

Introduction

Monitoring is the making of routine observations on health, productivity and environmental factors and the recording and transmission of these observations. Surveillance is a more intensive form of data recording than monitoring. It involves the continuous collection and interpretation of data collected during monitoring programs, usually with the recording of the identity of diseased individuals, with a view to detecting changes in a population's health. (1)

The Styrian Salmonella surveillance program in pigs includes breeding and fattening farms, slaughterhouses and meat processing plants. According to a two-way stratified cluster sampling the prevalence of Salmonella in pork was estimated in the initial phase. In the routine phase the surveillance of Salmonella is performed.

In this paper the sampling schemes (initial and routine phase) for the data collection in slaughterhouses are presented where major attention is paid on the influence of diagnostic tests calculating sample sizes and estimating Salmonella prevalence in pork.

Sampling scheme for the initial phase

The required sample size n (Equation 1) to estimate the prevalence \( p \) of Salmonella-positive pigs depends on the accuracy \( a \), type-one error \( \alpha \), the population size \( N \) and the prevalence \( p \). An estimator for a two-sided confidence limit \([p_1; p_2]\) for \( p \) is given in Equation 2, where \( F \) denotes the quantiles of a F-distribution and \( m \) the number of test-positive pigs.(2)

\[
n = \frac{\frac{\mu_1^2}{\alpha/2} p(1-p)}{\frac{\mu_1^2}{\alpha/2} p(1-p) + \frac{1}{N}}
\]

Equation 1:

\[
p_1 = \frac{m F_{2m,2(n-m+1)\alpha/2}}{n-m+1 + m F_{2m,2(n-m+1)\alpha/2}}
\]

Equation 2:

\[
p_2 = \frac{(m+1) F_{2(m+1),2(n-m-1)\alpha/2}}{n-m+(m+1) + F_{2(m+1),2(n-m-1)\alpha/2}}
\]

For a type-one error \( a \) of 0.05, \( \alpha \) population size \( N \) of 1.2 Mio. Styrian pigs, a desired accuracy \( a \) of 3% and an estimated prevalence \( p \) of 5% the required sample size \( n \) (equation 2) is 202.

The selection of the pigs for the Salmonella program is done according to a two-stage stratified cluster sampling scheme. At first stage the population is stratified with respect to geographical aspects and production areas. At the second stage the slaughterhouses, which are representing clusters in a statistical sense, are stratified into three strata with respect to their slaughter capacity. Strata 1 represents a capacity greater than 150,000 slaughtered pigs a year, strata 2 contains slaughterhouses with a capacity between 10,000 and 150,000 slaughtered pigs a year and strata 3 denotes the small slaughterhouses (capacity between 1000 and 10,000 slaughtered pigs a year). Using a pps-procedure (probability proportional estimated size) 5 slaughterhouses are selected from strata 1, 3 from strata 2 and 2 from strata 3. As 96% of the whole population is slaughtered in these 10 slaughterhouses the loss of representativity can be neglected. The sample size \( n \) is distributed to the slaughterhouses proportional to their capacity. To select the pigs in each slaughterhouse a systematic sampling procedure is used.
Sampling scheme for the routine phase

At the routine phase the proportion of animals tested from a herd is known. Each herd is classified as *Salmonella-positive* (diseased) or *Salmonella-negative* (non-diseased) based on the number of reactors. If the number of reactors is greater than or equal one the herd is classified as *Salmonella-positive*.

To calculate the number of herds ($n$) that need to be tested to establish that Salmonella is present at a level lower than the specified minimum prevalence ($d/N$) with a type-one error $\alpha$ the following approximate formula (3) can be used:

$$n \sim \left[ 1 - (1-\alpha)^{1/d} \right] \times (N - 1) + 1$$

For a type-one error $\alpha=0.05$, a population of $N=40,000$ fattening farms and a maximum prevalence of 1% equation 3 gives the sample size $n$ of 300 herds. The strategy to select the slaughterhouses where the herds shall be tested is the same like in the initial phase.

The number of animals to be tested from each herd can be calculated with the type-two error $\beta$ (2). The value $\beta$ is the probability of accepting the null hypothesis in the absence of disease and it is used as the threshold for the maximum number of reactors if the disease is not present. An estimator for the $\beta$ is given in Equation 4, where $N$ denotes the size of the herd, $M$ the expected number of infected animals from a herd and $n_b$ the number of animals to be tested.

$$\beta = \frac{\left( N - M \right)}{n_b^n(n-N)}$$

Equation 5:

$$\pi = \frac{\hat{p} + s_p - 1}{s_c + s_p - 1}$$

To select the animals of each herd a simple random sampling is used. If disease is present in a herd (cutoff number of reactors of 1), a rate of 20% is proposed, i.e. the expected number of infected animals is $M=0.2N$. For herds with more than 50 pigs and the type two-error $\beta$ of 0.05 a sample size of 12 pigs is required to establish whether the herd is free from disease or not. To be confident with a probability of 95% that the herd is free of Salmonella, all tested animals have to be *Salmonella-negative*.

The Influence of a diagnostic test

To estimate the proportion $\hat{p}$ of Salmonella-positive (Salmonella negative) animals (herds) diagnostic tests are used. Till now the calculations have been based on the assumption that the test used was perfect, i.e. sensitivity $= specificity = 1$. An estimation $\pi$ of the true prevalence $p$ with known sensitivity $s_c$ and known specificity $s_p$ is given by the following formula (4) where $\hat{p}$ is the observed proportion of animals with a positive test result.

To determine the required sample size for the monitoring program taking into account the sensitivity $s_c$ and specificity $s_p$ Equation 6 (5) can be used.

$$P(T^x = x) = \sum_{y=0}^{d} \binom{N}{y} \binom{N - d}{n - y} \prod_{j=0}^{\min(x,y)} \left( \frac{y}{j} \right) S_{c}^{y} (1-S_{c})^{y-j} \left( \frac{n-y}{x-j} \right) (1-S_{p})^{x-j} S_{p}^{n-x+y-j}$$

The required sample sizes ($n$) assuming $\alpha=\beta=5\%$, a herd size of $N=50$ and a supposed herd prevalence of 20% for selected values of sensitivity and specificity is given in the following table:
Table 1: Different sample sizes assuming sensitivity and specificity

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
<th>n</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>100%</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>95%</td>
<td>100%</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>100%</td>
<td>95%</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>95%</td>
<td>95%</td>
<td>28</td>
<td>3</td>
</tr>
</tbody>
</table>

Interpretation of table 1:

(i) For the case that the herd is Salmonella-negative, the probability of drawing less than c reactors from a sample of n animals is 1-a. (here 95%).

(ii) Is the herd infected with minimum prevalence of 20%, the probability of observing less than c reactors is a (here 5%).

(iii) It can be seen that the sample size increases if the sensitivity or specificity decreases.

References