

Herd-level interpretation of test results for epidemiologic studies of animal diseases

Jette Christensen^{a,*}, Ian A. Gardner^b

^aDanish Veterinary Laboratory, Bülowssvej 27, DK-1790 Copenhagen V, Denmark

^bDepartment of Medicine and Epidemiology, School of Veterinary Medicine,
University of California, Davis, CA 95616-8737, USA

Abstract

Correct classification of the true status of herds is an important component of epidemiologic studies and animal disease-control programs. We review theoretical aspects of herd-level testing through consideration of test performance (herd-level sensitivity, specificity and predictive values), the factors affecting these estimates, and available software for calculations. We present new aspects and considerations concerning the effect of precision and bias in estimation of individual-test performance on herd-test performance and suggest methods (pooled testing, targeted sampling of subpopulations with higher prevalence, and use of combinations of tests) to improve herd-level sensitivity when the expected within-herd prevalence is low. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Herd test; Bias; Pooled testing; Test combinations; Sensitivity; Specificity; Predictive values

1. Introduction

1.1. Concept and reasons for herd testing

Correct classification of herd status with respect to one or more pathogens is important in specific-pathogen-free (SPF) and other health-certification schemes, in risk-factor studies, in risk assessments of disease introduction, for diagnostic purposes and in disease-control programs. In these situations, the need for intervention usually is

* Corresponding author. Tel.: +45-35-300100; fax: +45-35-300120.

E-mail address: jc@svs.dk (J. Christensen)

determined by the status of the herd rather than by the status of each individual within the herd. The concept of a “herd test” and factors affecting herd-test results have been described by Martin et al. (1992) and Donald et al. (1994) and applied in studies reported in this journal and elsewhere (Carpenter and Gardner, 1996; Jordan, 1996; Jordan and McEwen, 1998).

With emerging animal-health programs aided by regional or national databases, more veterinary register data are now available for epidemiological analyses (Griffin et al., 1996; Losinger et al., 1997; Mousing et al., 1997; Carstensen and Christensen, 1998; King et al., 1998). These analyses typically use the herd as the unit of interest and therefore one issue is how to define the herd status for a large number of herds — in national registers, usually there are thousands of herds.

1.2. Definition of a herd

For purposes of this paper, we use the term “herd” to mean any cluster or aggregate of animals, e.g. litter, pen, paddock, farm, flock, barn, tank, pond, cattery, or a group of animals in a quarantine station. The location of animals that constitute the herd can be an important part of our definition of a herd. Therefore, the individual sites of a multi-site production or the whole aggregate of sites can be defined as a herd depending on the study in question. In some studies, it might be important to assume that the risk of disease is similar for all individuals within the herd relative to the risk among herds. In human populations, there are analogous clusters (households, villages, communities, child-care centres) which are also relevant to the discussion that follows. Therefore, the issues discussed in this paper may also have implications in human epidemiology.

1.3. Definition and objectives of a herd test

We define a herd test as an evaluation of a sample of (or all) animals from a herd and the application of decision rules that classify the herd as positive or negative based on the test results from individual animals. We use the generic terms “positive” and “negative”, to encompass the range of possible interpretations of herd status that are made by investigators (e.g. diseased or non-diseased, infected or non-infected, seropositive or seronegative, exposed or non-exposed, and immune or non-immune). To qualify as a herd test, the sample must include at least two animals from the herd where results are obtained by individual-animal testing or by pooled testing of a sample from two or more animals. Extreme examples of herd tests are when one pooled sample constitutes the herd test (e.g. a bulk-milk sample from all lactating cows) and when all animals are tested (when declaring freedom from disease) — but the herd test will usually be based on a sample of n animals (typically 10–20, although 5–60 are also used).

Decision rules that are applied to interpretation of a herd-test result usually will be applied at an individual-animal level first and then will be aggregated across results for all sampled animals from the herd. For a quantitative test such as an enzyme-linked immunosorbent assay (ELISA), it is therefore important to note that the herd-test result is a function of two decision thresholds: (1) the selected cut-off for the individual-test result (usually an optical density (OD) value or the OD value expressed as a percentage of the

OD of the positive reference sample, see Greiner and Gardner, 2000a for details), used to designate a sample as positive or negative, and (2) the selected cut-off value for herd-level interpretation (herd cut-off value). The latter value is the number (or proportion) of individual-animal positive test results that is sufficient to designate the herd as positive. For example, if a herd cut-off of 2 is used, then ≥ 2 positive results in n samples would be interpreted as a positive herd-test result and 0 or 1 positive samples would be a negative herd-test result (Appendix A).

The objective of the herd test is — similar to the objective of an individual-animal test — to classify herds as positive or negative (usually with a prescribed level of confidence). When the reason for the classification is diagnostic or declaration of an infection status, it is important to define a status for all herds. However, in some epidemiological studies it may be more important to minimise misclassification of true herd status rather than have a definitive status on all herds. Therefore, the herd classification could have three categories: positive, inconclusive/unknown, or negative. Also, in some epidemiological studies and control programs, a classification reflecting a level of infection in the herd may be preferred; e.g. in the Danish Salmonella Control Program, herds are assigned to Salmonella level 1, 2, or 3 for low, moderate, or high sero-prevalence, respectively (Mousing et al., 1997).

Based on our definition, the following examples would qualify as herd tests to classify the herd as positive or negative:

- serologic testing of 20 pigs in a herd for *Mycoplasma hyopneumoniae*,
- bacteriological culture of composite quarter-milk samples from cows (equal volumes of milk from each quarter pooled into a single sample per cow) for mastitis pathogens,
- bulk-milk tank test for an infectious agent (e.g. bovine leukemia virus, bovine virus diarrhoea virus) or for a residue detection test for beta-lactam antibiotics,
- pooled faecal sample from five pigs in each of 10 pens for salmonella culture,
- samples of caudal kidney from 60 fish pooled for testing for infectious pancreatic necrosis virus.

A test of tanker-truck milk from multiple herds, a composite sample of milk from four quarters of a single cow, and faecal samples from five pigs in five different herds pooled into a single sample would not meet the definition of a herd test.

Usually, interpretation of herd-test results is more complicated than interpretation of individual-test results because two decision thresholds are required and herd-test results are influenced by multiple factors (see Section 2.2). For some diseases, the herd status might be easier to establish than an individual's status — especially when test sensitivity for individual animals is low but the procedure usually is perfectly specific (e.g. faecal culture for subclinical salmonella infection in pigs).

In this paper, we present theoretical aspects of herd-level testing through a review of herd-test performance (herd sensitivity, specificity and predictive values), the factors affecting these estimates, available software for calculations, and pooled testing. In addition, we present new aspects and considerations concerning the effect of precision and bias in estimation of individual-test performance on herd test performance and suggest methods (pooled testing, targeted sampling of subpopulations with higher prevalence, and use of combinations of tests) to improve herd sensitivity.

2. Review

2.1. Herd sensitivity and herd specificity

Herd-level sensitivity (HSe) is the probability that a positive (diseased, infected, seropositive, exposed or immune) herd yields a positive herd-test result and herd-level specificity (HSp) is the probability that a negative (non-diseased, non-infected, seronegative, non-exposed or non-immune) herd yields a negative herd-test result (Martin et al., 1992). The false-negative and false-positive herd proportions (rates) can be calculated from the respective HSe and HSp values.

The gold standard for a herd's true status might be determined by a single test or a combination of tests. We note that true herd status should be defined based on the status of all animals in the herd rather than the sample because it is possible to have a sample that contains no truly positive animals even though the herd is truly positive.

For an infectious disease, non-infected herd status is often determined by combinations of ongoing negative results of necropsies, clinical examinations, and laboratory tests, and knowledge that the herd uses a defined biosecurity program and is closed to introductions of animals. Once a herd is designated as non-infected, all individuals in the herd automatically assume the same status. Determination of infected herd status can be based on similar tests and knowledge of likely transmission patterns and persistence of the pathogen of interest. However, in contrast to the situation for non-infected herds, the infection status of each animal is not known automatically because some animals will not be infected. This might occur because infection is transient, or because host susceptibility varies with age, sex, or immune status. Furthermore, management and environmental factors probably affect within-herd transmission of many endemic infectious agents. These effects are manifested as herd prevalences of <1 and among-herd variation in the proportion of infected animals.

An infected herd could be designated as such because some infected animals reacted to the test (reflecting the sensitivity (Se) of the individual-test) or because some non-infected animals reacted to the test (reflecting the imperfect specificity (Sp) of the individual-test). Hence, false-positive individual-test results could correctly specify that the herd is infected even though the infected animals in the sample tested negative.

2.2. Factors affecting herd sensitivity and herd specificity

The HSe and HSp are dependent on the Se and Sp of the individual-test, the number of animals tested (n), the true within-herd prevalence (TP) in infected herds, the herd cut-off value (e.g. 1, 2, or 3 positive test results) used to classify the herd as positive (Martin et al., 1992), and the variation in Se, Sp and TP among herds (Donald et al., 1994).

For all calculations and examples that follow, we assume that the sample size for the herd test is small compared with the herd size and that binomial probabilities are appropriate. Hypergeometric probabilities can be readily used as an alternative to the binomial for computer calculations (see Section 2.5).

Martin et al. (1992) showed that the link between the herd- and individual-test performances can be developed by considering the linear relationship between apparent

within-herd prevalence (AP) based on the results of an imperfect test (T), TP, and the bias between the two values caused by imperfect Se and Sp,

$$AP = \Pr(T+) = Se \times TP + (1 - Sp) \times (1 - TP) \quad (1)$$

For a non-infected herd (TP=0), Pr(T−) or 1−AP can be obtained by substitution of TP=0 into Eq. (1) and equals Sp. Assuming that Sp is a constant value for all animals in the herd, the probability that all animals will test negative (HSp) is Sp^n for a herd cut-off value of 1. Therefore, the probability of finding at least one positive animal (false-positive herd proportion) is $1 - Sp^n$. For an infected herd (TP>0), the probability that a randomly selected animal will test negative (i.e. Pr(T−)=1−AP) is a function of Se, Sp and TP as shown in Eq. (1). Assuming that Se and Sp are constant for all animals in the herd, then the probability of finding zero positive individuals (false-negative herd proportion) is $(1 - AP)^n$. Therefore, HSe for a herd cut-off of 1 positive to class the herd as positive is $1 - (1 - AP)^n$.

Formulas for HSe and HSp as a function of cut-off values >1 are complicated but can be readily generated from the binomial (Martin et al., 1992) or hypergeometric distribution, if necessary. Also, it is possible to consider the proportion positive rather than the number positive — especially if the number tested differs substantially among herds.

For the remainder of the paper, we consider two hypothetical tests: bacteriologic culture with Se=0.5 and Sp=0.999 (e.g. culture of faeces for subclinical salmonella infection in pigs) and a serologic test with Se=Sp=0.9 (e.g. ELISA for *Mycoplasma hyopneumoniae* in pigs) to demonstrate the effects of Se and Sp on HSe and HSp. Mostly we restrict calculations and examples to a herd cut-off value of 1 positive-test result for designating a herd as positive. We chose this value because it is commonly used for definition of true herd status when bacteriologic and virologic methods (assumed to have Sp >0.999) are used; also, when serologic methods are used, use of higher thresholds (2, 3, etc.) is still uncommon.

We comment on some relationships among factors affecting HSe and HSp.

- When testing a fixed number of animals, HSe increases as within-herd AP (and hence TP) increases (Fig. 1).
- As the number of animals tested increases, HSe increases. An increase in sample numbers from 10 to 20 or 20 to 60 for a test with Se=0.5 and Sp=0.999 produces substantial increases in HSe if within-herd AP is low to moderate (Fig. 1). For this example, once AP is >0.4, then changes in sample size have smaller effects.
- In a non-infected herd, as the number of animals tested with an imperfectly specific test increases, the probability of having at least one false-positive test result increases — resulting in a lower HSp. Reductions in HSp are less as the individual-level test becomes more specific (Fig. 2).
- As the number of animals used to classify the herd as positive (herd cut-off value) is increased, there is a corresponding increase in HSp with a decrease in HSe.

Donald et al. (1994) showed that correlations between test results from different animals in the same herd (sensitivity and specificity correlations) and variation in TP

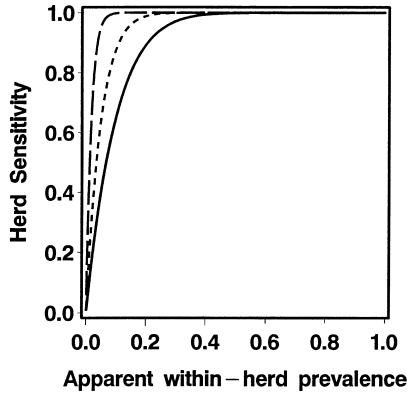


Fig. 1. The relationships among herd sensitivity ($1-(1-AP)^n$), apparent within-herd prevalence (AP), and number of samples for a herd test based on tests with individual $Se=0.5$ and $Sp=0.999$. Herd-test cut-off=1.

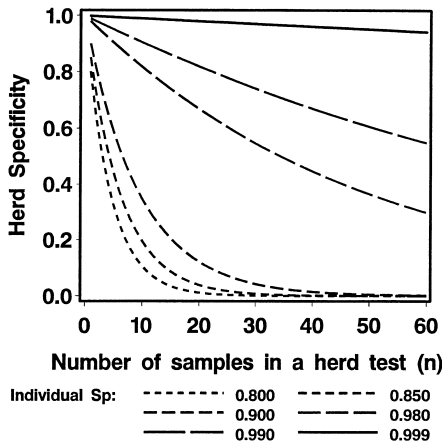


Fig. 2. The relationships among herd-level specificity (Sp^n), individual-level specificity (Sp), and number of samples in a herd test for different individual-test specificities. Herd-test cut-off=1.

among herds (disease correlation) also affect these relationships. A plausible biologic explanation for the occurrence of sensitivity correlations is that Se varies with stage and severity of infection (Ransohoff and Feinstein, 1978). Hence, Se could vary among herds since some herds might have been recently infected, others might be chronically infected, and in others, the owner/manager might remove or treat clinically affected animals. A specificity correlation might exist because of herd-level variation in use of vaccines that induce antibodies that cross-react on serologic tests and the occurrence of cross-reacting organisms which cluster by herd. The effects of correlations identified by Donald et al. (1994) are complex but include (for a sample size of 10 and $Se=Sp=0.9$):

- An increase in the sensitivity correlation reduces HSe slightly at a fixed herd cut-off value but there is no effect on HSp.

- An increase in the specificity correlation decreases HSe slightly at lower herd cut-off values (1 or 2) but changes are minimal at higher herd cut-off values (>2).
- The range of HSp is wider with low specificity correlations than with higher specificity correlations.
- A large disease correlation diminishes HSe but there is no effect on HSp.

In addition to the factors described above, a herd test of a fixed size for an animal pathogen might have several HSe values depending on the stage of infection in the herd and timing of the herd test compared with determination of true herd status.

2.3. Herd predictive values

Herd-level predictive values are the herd-level analogues of individual-test predictive values. They depend on all the factors affecting HSe and HSp, and on the true proportion of positive herds (HTP) (Martin et al., 1992). In this context, HTP means the pre-test probability that a herd is positive. This probability could vary by geographic location, herd type, herd size, or other risk factors.

$$\text{HPV}_+ = \frac{\text{HSe} \times \text{HTP}}{\text{HSe} \times \text{HTP} + (1 - \text{HTP}) \times (1 - \text{HSp})} \quad (2)$$

$$\text{HPV}_- = \frac{(1 - \text{HTP}) \times \text{HSp}}{(1 - \text{HTP}) \times \text{HSp} + \text{HTP} \times (1 - \text{HSe})} \quad (3)$$

As for individual-test interpretation, HPV_+ is related directly to the HTP whereas the HPV_- is related inversely to HTP.

2.4. Sampling and sample size considerations

To estimate the required number of samples (n) to detect at least one truly positive animal in a herd, two values are necessary: the required level of confidence (C) or desired HSe (often at least 0.95), and the within-herd prevalence of truly positive animals (TP) (Martin et al., 1992):

$$n \geq \frac{\log(1 - C)}{\log(1 - \text{TP})} \quad (4)$$

Eq. (4) can be modified to include the Se and Sp of the test. First, TP is replaced by AP and therefore, the new interpretation of n is “the sample size needed to detect at least one test positive animal”. Next, AP is replaced by the right-hand side of Eq. (1) to obtain the general formulation for the sample size needed to detect at least one test positive animal.

$$n \geq \frac{\log(1 - C)}{\log(\text{Sp}(1 - \text{TP}) + (1 - \text{Se})\text{TP})} \quad (5)$$

with the special cases:

- If $\text{Se}=1$ and $\text{Sp}=1$ then $\text{AP}=\text{TP}$, hence $n \geq \log(1 - C) / \log(1 - \text{TP})$.
- If $\text{Se}<1$ and $\text{Sp}=1$ then $n \geq \log(1 - C) / \log(1 - \text{Se} \times \text{TP})$.
- If $\text{Se}=1$ and $\text{Sp}<1$ then $n \geq \log(1 - C) / \log(\text{Sp}(1 - \text{TP}))$

It is clear that as n increases: (1) for a given Se, Sp, TP and herd cut-off value, HSe increases and HSp decreases; (2) the herd apparent prevalence or proportion of test positive herds (HAP) increases and HPV+ decreases (Martin et al., 1992).

For two-stage cluster surveys to substantiate disease freedom of a county or region (with many herds), special sample size calculations are needed to incorporate the sampling design, test performance and the probability of type-I and type-II errors. Analysis of data from these surveys will determine whether disease is present or absent, with a specified probability (Cameron and Baldock, 1998a,b). The computations needed for the exact sample size calculations considering imperfect tests and a finite population are extensive and algorithms have been developed and implemented in the computer program *FREECALC* (Cameron and Baldock, 1998a,b), which is described in Section 2.5. As shown in the example given by Cameron and Baldock (1998a) for Johne's disease, use of imperfect tests (Se=0.6 and Sp=0.99) in finite populations (herd size=265) may make it virtually impossible to distinguish between infected and non-infected status with high confidence when within-herd TP is low (0.02).

2.5. Computer software

Three shareware programs are available to facilitate calculation of HSe and HSp values and guide choices of sample size and of the herd cut-off value for determination of herd status. The programs were designed for different reasons but for practical purposes yield the same results with a few exceptions. All programs assume random sampling, that the individual-level Se and Sp values are known with certainty and are unbiased. Two of the three programs assume that Se, Sp and TP are constant from herd to herd. Because precision and bias (see Sections 3.1 and 3.2) can have substantial effects on the estimates and the confidence intervals (CI), we recommend that the numeric values estimated by the programs be used as a guide only.

HERDACC version 3 is Windows-based (available at <http://epiweb.massey.ac.nz/>) and generates a series of output tables of HSe and HSp for various combinations of inputs (up to five prevalences and various sample sizes and herd cut-off values) in a single run (Jordan, 1996). The investigator has two choices — binomial sampling (with replacement) or hypergeometric (without replacement). For practical purposes, the estimates are similar until the sample sizes exceed about 20% of the population size. For hypergeometric calculations, *HERDACC* uses an approximation based on the expected number of test positive animals in the population, which must be an integer value — note that this rounding may cause differences from values estimated in *FREECALC*. One application of the program allows determination of two herd cut-off values that allow prescribed values of HSe and HSp to be obtained. This approach has been applied to Johne's disease in cattle (Jordan, 1996). A new software program which includes Monte Carlo simulation has been developed by the same author to allow for the possibility of among-herd disease correlation, uncertainty in estimates of Se and Sp, and more complex sampling protocols (Jordan and McEwen, 1998). The program is likely to be more widely available in the future.

FREECALC is an epidemiological calculator for the design and analysis of surveys to detect disease or to substantiate freedom from disease (Cameron and Baldock, 1998a,b).

Windows and DOS versions of the program can be downloaded from <http://epiweb.massey.ac.nz/> or from <http://www.pnc.com.au/~angus>. There are two major differences from HERDACC. First, the program offers three choices for calculations: binomial, exact hypergeometric and an approximation. Second, the program uses a different interpretation of the herd cut-off value. The cut-off value is the maximum number of test positive animals that a herd can have and still be considered free of disease. Hence, the cut-off in FREECALC is always one less than the value estimated in HERDACC. Formulas, theory and sample calculations are shown in Cameron and Baldock (1998a,b).

AGG.EXE (available at http://www1.vetmed.fu-berlin.de/~mgreiner/validation_course/announce.htm) is a Fortran-compiled DOS-based program that was written to examine the theoretical effects of correlation between tests and within-herd TP on HSe and HSp (Donald et al., 1994). The program extends the calculations of Martin et al. (1992) to include betabinomial distributions to model variation in Se, Sp and TP among herds. The output is a series of HSe and HSp values for the different herd cut-off values. When the correlations are set to zero, the program yields the same output as HERDACC when the latter is used with the binomial-sampling option.

2.6. Use of pooled samples as herd tests

In Sections 2.1–2.4, we considered testing of individual samples for determination of herd infection status. However, pooled testing of milk, feces, eggs and animal tissues is being increasingly used as a cost-effective alternative to testing of individual samples. Mostly, pooled testing is used when individual results are not needed or when samples can only be obtained in a pooled form. A herd test might involve testing of single or multiple pools from a herd. If a single pooled sample is collected per herd, the only question that can be answered is whether the herd is infected or not. An estimate of within-herd TP requires at least two pools/herd and we refer interested readers to Cowling et al. (1999) for a complete description of maximum likelihood and Bayesian approaches to prevalence estimation based on pooled sampling.

The primary advantage of a pooled test over an individual-test is that more individuals can be represented in pooled tests for the same fixed laboratory cost. This sample size effect can increase HSe (and reduce HSp) and we explore this effect formally in Section 3.4. The advantage of pooling versus individual testing is greatest when prevalence is low (<0.05) and decreases as prevalence increases. One disadvantage of pooled testing is logistical constraints associated with processing of larger sample weights or volumes in the laboratory. Also, there is a potential decrease in sensitivity compared with individual-animal testing.

We define the pool sensitivity (PSe) as $\Pr(\text{pool } T+ \mid \text{pool } D+)$ and pool specificity (PSp) as $\Pr(\text{pool } T- \mid \text{pool } D-)$. Factors affecting PSe are probably complex and have not been extensively considered in veterinary medicine. We comment briefly on the likely effects of dilution, concentration of analyte (e.g. antibodies) and sampling probabilities on the PSe of a bulk-milk antibody detection assay that is used when within-herd TP is low. Consider a 100-cow herd with one infected cow where the pooled test involves 10 milk pools with 10 cows per pool or a single pooled test of milk from 100 cows.

Assuming a fixed milk concentration of antibody for the infected cow, the likelihood of a positive test result probably will be lower in the second scenario. The dilution effect on PSe will be dependent on the infected cow's concentration of antibody in relation to the selected cut-off value for the assay and her milk production compared with other cows in the pooled sample. However, if the infected cow is dry, has mastitis or has been treated with antibiotics and is not contributing milk to the bulk tank, then the probability of inclusion in the sample will be zero (i.e. inclusion probability is dependent on infection status), and the pooled test result may be negative even though the herd is infected.

3. New aspects and considerations

3.1. Precision of herd sensitivity and specificity estimation

Formulas used by Martin et al. (1992) to calculate HSe and HSp from Se and Sp assumed that the latter were known with certainty. However, estimates of Se and Sp inevitably are based on evaluation of finite numbers of truly positive and truly negative animals. Often, 100–200 animals are used (Nielen et al., 1994; Acorda et al., 1995; Geishauser et al., 1998) and only rarely are >1000 samples used (Smitsaart et al., 1998; Vancini et al., 1998). Accordingly, it is important that uncertainty in individual-level Se and Sp estimates be incorporated into calculations of HSe and HSp. We therefore estimated 95% CI for individual Se and Sp with varying numbers of animals being used in the test-evaluation studies. We then used these estimates to calculate 95% CI for HSe and HSp by substitution of the confidence limits for AP and Sp, respectively, and herd-test sizes ($n=10$ or 20) into the formulas for HSe and HSp.

3.1.1. Herd sensitivity estimation

The precision of the HSe estimate is dependent on the precision of the individual-level Se and Sp estimates (assuming that TP is known). First, we calculated a 95% CI for AP following the approach of Rogan and Gladen (1978) for estimation of TP. Assuming no sampling variability (i.e. that the sample prevalence is equivalent to the herd prevalence), the variance estimate for AP is

$$\text{Var}(\widehat{\text{AP}}) \approx \text{TP}^2 \times \frac{\widehat{\text{Se}} \times (1 - \widehat{\text{Se}})}{N} + (1 - \text{TP})^2 \frac{\widehat{\text{Sp}} \times (1 - \widehat{\text{Sp}})}{M} \quad (6)$$

where the sensitivity estimate ($\widehat{\text{Se}}$) is estimated based on N truly positive animals and the specificity estimate ($\widehat{\text{Sp}}$) is estimated based on M truly negative animals. An approximate 95% CI for AP is

$$95\% \text{ CI for AP} \approx \widehat{\text{AP}} \pm 1.96 \sqrt{\text{var}(\widehat{\text{AP}})} \quad (7)$$

The 95% upper and lower limits for AP were used to obtain an approximate 95% CI for HSe.

For a test with $\text{Se}=\text{Sp}=0.9$, we evaluated the effects of using between 10 and 300 positive (N) and negative (M) animals in test-evaluation studies and then determined the

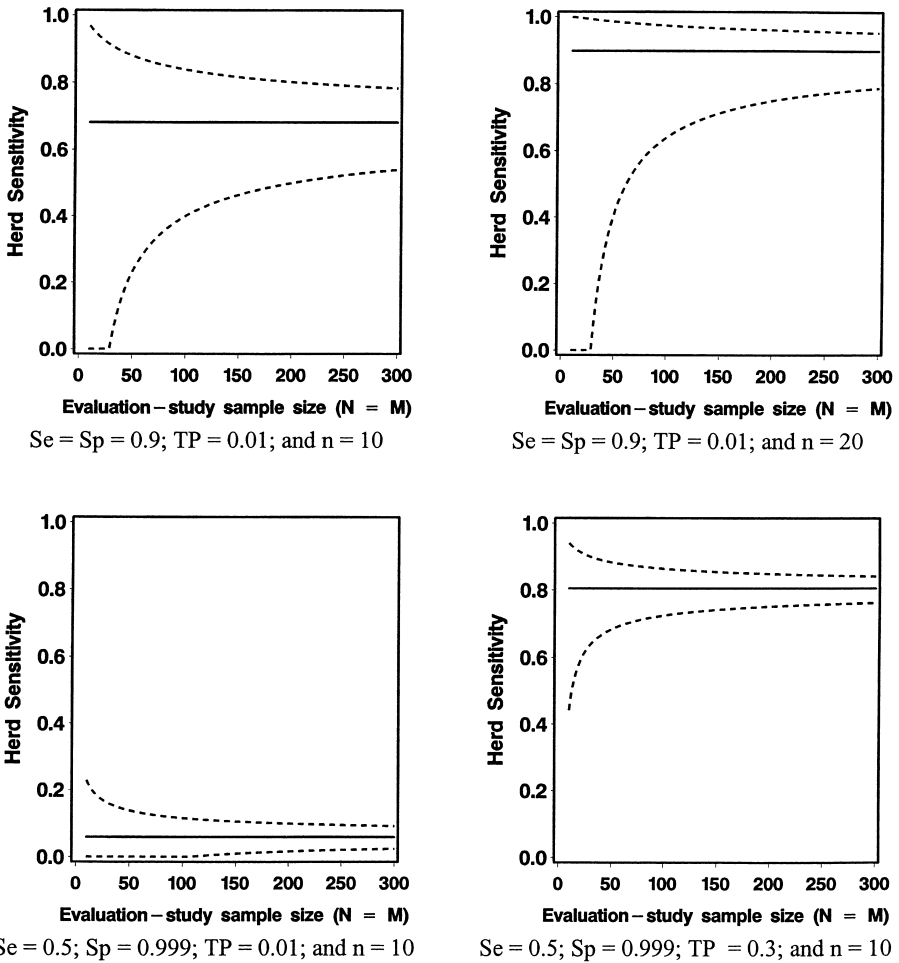


Fig. 3. Point (solid) and 95% CI (dashed lines) for herd-level sensitivity estimates for numbers in the evaluation study ($N=M$) ranging from 10 to 300, $n=10$ and 20 in herd tests, true prevalence 0.01 or 0.3 with the individual-test characteristics of $Se=Sp=0.9$ or $Se=0.5$ and $Sp=0.999$.

effects on HSe when $n=10$ or 20 animals were used for herd tests. Test-evaluation studies with small sample sizes and point estimates close to 0.5 had the widest CIs (Fig. 3). Once sample sizes reached 300, effects on width of the CI were minimal.

3.1.2. Herd specificity estimation

The precision of the HSp estimate depends only on the precision of Sp. For $Sp=0.9$ and 0.999, we examined the effect of evaluation-study sample sizes between $M=10$ and 300 truly negative animals and when a herd test of $n=10$ or 20 animals was done in a truly negative herd (Fig. 4). Evaluation studies of small size ($M<50$) resulted in wide 95% CIs for HSp and the width of the CI was greater when HSp point estimates were close to 0.5.

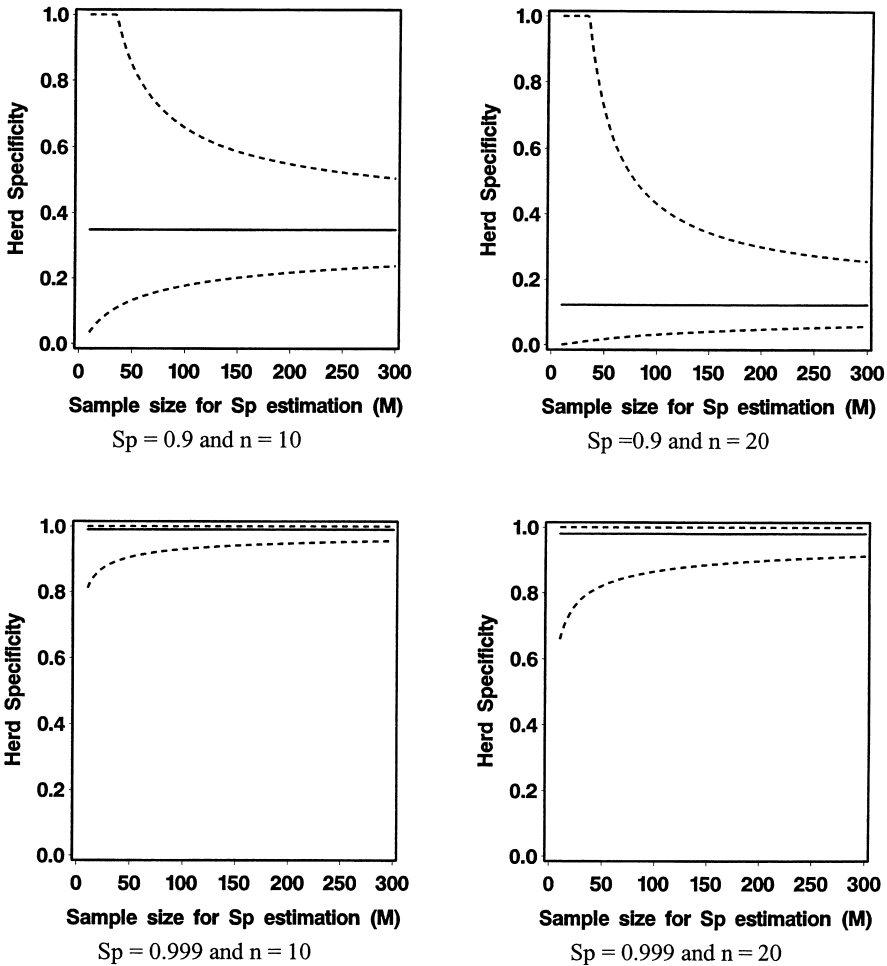


Fig. 4. Point (solid) and 95% CI (dashed lines) for herd-level specificity estimates for numbers in the evaluation study (M) ranging from 10 to 300, a herd test of $n=10$ and 20 with the individual-test characteristics of $Sp=0.9$ and $Sp=0.999$.

3.2. Bias and misclassification

3.2.1. Bias in herd sensitivity and specificity estimation

Overestimation of Se and Sp often occurs when estimates are derived from laboratory experiments, when an unrepresentative spectrum of field samples is used, or when other biases occur (Ransohoff and Feinstein, 1978; Carrington-Reid et al., 1995; Greiner and Gardner, 2000b). Also, Sp might be underestimated when apparently disease-free animals are tested in a known-infected population (Rogers et al., 1989).

The extrapolation from the individual-test result to the herd-test result (Sections 2.1–2.4) assumed that the estimate was unbiased. To demonstrate the effects of biased

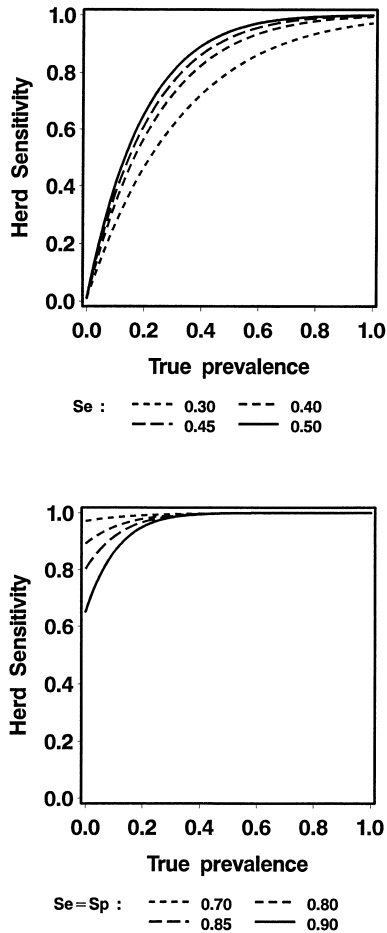


Fig. 5. The relationship between herd-level sensitivity and the true prevalence when estimates of individual-test sensitivity and specificity are biased. In the upper diagram, biased estimate of individual-test $Se=0.5$ and the estimate of $Sp=0.999$ is unbiased. In the lower diagram, the biased estimates are $Se=Sp=0.9$. The number of samples in both herd tests is $n=10$.

estimation of Se and Sp , we recalculated HSe and HSp for biases that we considered to be plausible.

When the estimated Se was 0.5 (biased compared with true values of $0.3, 0.4,$ and 0.45) and Sp was 0.999 (unbiased), the effect was values of HSe that were also too large over a wider range of TP values (Fig. 5, upper). For a test with (over)estimated $Se=Sp=0.9$, we calculated the HSe if the true individual Se and Sp were both lower ($Se=Sp=0.85, 0.8,$ and 0.7) for a range of TP values ($0-1$) and $n=10$ samples used in the herd test. Once TP was >0.4 , the overestimation had no effect on the HSe because HSe was almost 1 — but for lower TP , overestimation of both Se and Sp resulted in HSe values that were too low (Fig. 5, lower).

Biased estimates of individual Sp had major effects on HSp estimates for all values of n . For example, for $n=20$, biased estimates of 0.999 (too high) or 0.98 (too low) when the true value was 0.99 yielded HSp values of 0.98 (0.999^{20}) or 0.67 (0.98^{20}) compared with the true value of 0.82 (0.99^{20}).

3.2.2. Misclassification of herd disease status in veterinary-register data

Veterinary registers that are used for disease-control or eradication programs (DCP) involve data for many herds whose disease status changes over time. In these registers, the disease status of herds might be misclassified and this can be important when these data are used for epidemiologic studies. We discuss misclassification in DCP and give an example to clarify its possible effects.

Misclassification of disease status in a DCP is possible because the designated status depends both on a formal herd test (disease surveillance) and regulations concerning the herd's compliance with the program. Depending on the situation, the misclassification can be used as a tool in the DCP. For example, if herds fail to have regular testing done (e.g. an annual herd test), they will be classified as diseased and animal movements will be restricted. Some of these herds will not be infected (false positives) — but this might be a strong incentive to comply with the DCP.

Changes in disease definition over time (e.g. when the prevalence of disease decreases as a consequence of the control program) may cause the direction of the misclassification to be time-dependent. In the early stages of a DCP, herds might only be classified as diseased when a confirmed diagnosis has been made because there are insufficient resources to control the disease in all herds. In the final phase of an eradication program (when disease occurrence is sporadic and the economic consequence of a few cases is great), even a suspicious or false-positive individual-test result may cause the herd to be assigned to the diseased group (false positives).

If the register includes the results of individual-animal tests, one solution to minimise misclassification is to define a herd-level test (HT) and a herd cut-off value to define herds as positive or negative. Based on the decision rule, an HT classification is given to all herds in the register.

The following example illustrates these issues. Assume that 1000 herds are tested in two situations: (1) at the beginning of a DCP when prevalence is high ($TP=0.75$), and (2) in the final phase of the DCP when prevalence has decreased ($TP=0.02$). Initially, 20% of herds comply with the program and have a verified DCP-positive status — of these 175/200 are truly diseased. Eighty percent of herds (“unknown”, “non-compliant” and “not tested”) are classified as DCP-negative but their true status is latent (Fig. 6, upper left). At the end of the program, assume that the status of 5% of herds is unknown. Because of DCP requirements, these herds are automatically classified as infected (DCP-positive) (Fig. 6, upper right). When prevalence is low, note that the proportion of DCP-positive herds is about 3-fold higher than the proportion of truly infected herds. In contrast, when prevalence was high, the proportion of the DCP-positive herds was about 4-fold lower than the proportion of truly infected herds.

Now assume that herds are classified by an HT instead of the DCP status. When $TP=0.75$ (Fig. 6, lower left), we assume that 90% of the 800 DCP-negative herds have no or insufficient testing to assign an HT result — these herds are excluded non-

High prevalence (TP = 0.75)				Low prevalence (TP = 0.02)					
		True +	True -			True +	True -		
DCP	+	175	25	200	DCP	+	18	50	68
Status	-	575	225	800	Status	-	2	930	932
		750	250	1000			20	980	1000
60% ((575+25)/1000) of the herds are misclassified				5.2% ((2+50)/1000) of the herds are misclassified					

High prevalence (TP = 0.75)				Low prevalence (TP = 0.02)					
		True +	True -			True +	True -		
HT	+	175	25	200	HT	+	18	5	23
	-	58	22	80		-	2	930	932
		233	47	280			20	935	955
29.6% ((58+25)/280) of the herds are misclassified and there are missing herd-test results for 720 herds				0.7% ((2+5)/955) of the herds are misclassified and there are missing herd-test results for 45 herds					

Fig. 6. Classification of 1000 herds in a hypothetical disease-control program (DCP). At the beginning of the DCP, prevalence is high (TP=0.75) and in the final phase when prevalence has decreased, TP=0.02 (Upper). In the beginning (upper left), the information from 80% of the herds is incomplete (missing information or insufficient testing) but all those herds are categorised as DCP-negative. Among DCP-negative herds, the distribution of True+ and True- results is latent. At the end of the DCP (upper right), 5% of the herds do not comply with the DCP; therefore, they are categorised as DCP-status positive although they are truly negative. In the lower row, a formal herd test (HT) is used instead of DCP-status and herds with no recent formal tests are excluded.

differentially with regard to true disease status. Similarly, when TP=0.02 (Fig. 6, lower right), 90% of the 50 DCP-positive but non-infected herds will be excluded because there are insufficient data for an HT result. In both situations, use of the HT result reduces the proportion of misclassified herds and improves HSe and HSp when prevalence is high and low, respectively.

Next we assume that two outcome definitions (DCP and HT) are used in a risk-factor study when TP=0.02. Further assume that 50% of the truly positive and 25% of the truly negative herds were exposed to a risk-factor (*E*) and that there is no misclassification of exposure. The true unknown odds ratio (OR) is 3 and use of HT rather than DCP status results in a less-biased OR estimate (Fig. 7, 2.49 compared with 1.36, respectively).

Clearly, if the objective was to assess risk factors for herd infection, the use of DCP status would fail to confirm the exposure as a risk-factor. Hence, it is preferable to define a herd infection status for the study by combining individual-test results into an HT result. Theoretically, it should also be possible to adjust OR estimates in a risk-factor study if the HSe and HSp of the test used to determine the outcome were known (see Greiner and Gardner, 2000a for a description of approaches).

Low prevalence (TP = 0.02)				
Disease classified by true status:		True +	True -	
OR = 3.0 (1.11 - 8.13) ^a ; P = 0.018 ^b	E +	10	245	255
Conclusion: E+ is a risk factor for disease (True +)	E -	10	735	745
		20	800	1000

Disease classified by DCP status:		DCP+	DCP-	
OR = 1.36 (0.76 - 2.39); P = 0.249	E +	21	234	255
Conclusion: No significant association between E+ and DCP +	E -	46	699	745
		67	933	1000

Disease classified by herd-test (HT) status:		HT +	HT-	
OR = 2.49 (0.95 - 6.37); P = 0.045	E +	10	234	244
Conclusion: E+ is a risk factor for HT+	E -	12	699	711
		22	933	955

^a Exact 95% CI; ^b Fisher's exact test (two tailed)

Fig. 7. A hypothetical risk-factor study where TP=0.02; 50% of the truly positive and 25% of the truly negative herds are exposed to a risk-factor (E). For the DCP-status, HSe=0.9 and HSp=0.95 and for the HT, HSe=0.9 and HSp=0.99 (Fig. 6).

3.3. Estimation of herd sensitivity and herd specificity from field studies

Lack of data on individual Se and Sp or substantial variation in TP among herds may preclude theoretical estimation of HSe and HSp. However, direct evaluation of HSe and HSp from field studies is possible and should provide more-realistic estimates of the parameters. Evaluation of HSe and HSp from field data avoids the need to extrapolate individual-test results to the herd when the individual-test results are based on experimental infections and makes no assumptions about a constant within-herd TP. However, the problem of defining a gold standard for true herd status and avoiding biased evaluation of the herd test needs careful consideration (see Greiner and Gardner, 2000b for a description of bias-reduction strategies).

3.3.1. Example of direct estimation of herd sensitivity and herd specificity

Some ELISA tests have been introduced as herd tests rather than as individual-animal tests (Hoorfar and Bitsch, 1995). For these tests, the individual-animal Se and Sp values are of minor importance. For example, the herd-test performance of ELISA tests used for monitoring of respiratory diseases in pigs — *Mycoplasma hyopneumoniae* and *Actinoba-*

cillus pleuropneumoniae serotype 2 — has been estimated using data from the Danish SPF scheme (Sørensen et al., 1992, 1993, 1997). The following data were obtained for *M. hyopneumoniae* in 121 SPF pig herds (Sørensen et al., 1992) based on newly detected infections using a herd test of 20 pigs with at least one positive ELISA result considered to be evidence of infection. The gold standard evaluation of herd *M. hyopneumoniae* status was based on clinical examinations, and on pathological and cultural examination of lungs from slaughter pigs. The HSe and HSp and the exact binomial CI were calculated: HSe=14/15=0.933 (95% CI=0.681–0.998) and HSp=102/106=0.962 (95% CI=0.906–0.990).

3.4. Estimation of herd sensitivity and herd specificity based on pooled tests

In this section, we describe the theoretical relationship between individual-animal and pooled test performances, and consider the herd sensitivity and specificity when multiple pools are used in the herd test. We show that pooling samples can increase HSe when the expected TP is low — but at the cost of reduced HSp (which may be substantial in some instances).

We motivate the discussion with a simple example when pool size is small and where it is reasonable to assume comparable performance of individual and pooled tests. Consider a pooled faecal-culture method for subclinical *Salmonella* in pigs that has PSe=Se=0.5 and PSp=Sp=0.999. The study herd has TP=0.05 and a herd test is defined by four pools of five (20 animals) instead of four individual tests. Substitution of these values (change in n from 4 to 20) in the formulas for HSe and HSp (see Section 2.2) indicates that HSe when based on pooled samples would theoretically increase from 0.10 to 0.41 and HSp would decrease from 0.996 to 0.98.

If only estimates of Se and Sp (and not PSe and PSp) are available, but the herd test is based on pooled tests, estimation of HSe and HSp is more complex because we have to make assumptions about PSe and PSp. We could assume that PSe=Se and PSp=Sp, as in the example above. However, it is likely that the PSe will be lower than Se — especially when TP is low and pool size is large. Conversely, PSp should exceed Sp because dilution should make it less likely to have a false-positive pooled test result than a false-positive individual-test result.

Herd testing based on pooled samples involves two levels of aggregation: pool and herd. We note again that a herd test could be based on single or multiple pools per herd. We assume in the following discussion that ≥ 1 D+ (or T+) animals in a pool of size k qualifies the pool as D+ (or T+) and that ≥ 1 D+ (or T+) pools in the herd qualifies the herd as D+ (or T+). The proportion of D+ (or T+) animals is not directly observable.

Hereafter, we use HPSe and HPSp to denote herd sensitivity and specificity estimates, respectively, that are based on pooled rather than individual tests. For a herd test involving r pools from an infected herd, and assuming no clustering of TP, PSe and PSp.

$$\begin{aligned}
 \text{HPSe} &= \Pr(\geq 1 \text{ pool tests positive} \mid \text{TP} > 0) \\
 &= 1 - \Pr(r \text{ pools test negative} \mid \text{TP} > 0) \\
 &= 1 - \Pr(\text{each pool tests negative} \mid \text{TP} > 0)^r \\
 &= 1 - [(1 - (1 - \text{TP})^k) \times (1 - \text{PSe}) + (1 - \text{TP})^k \times \text{PSp}]^r
 \end{aligned} \tag{8}$$

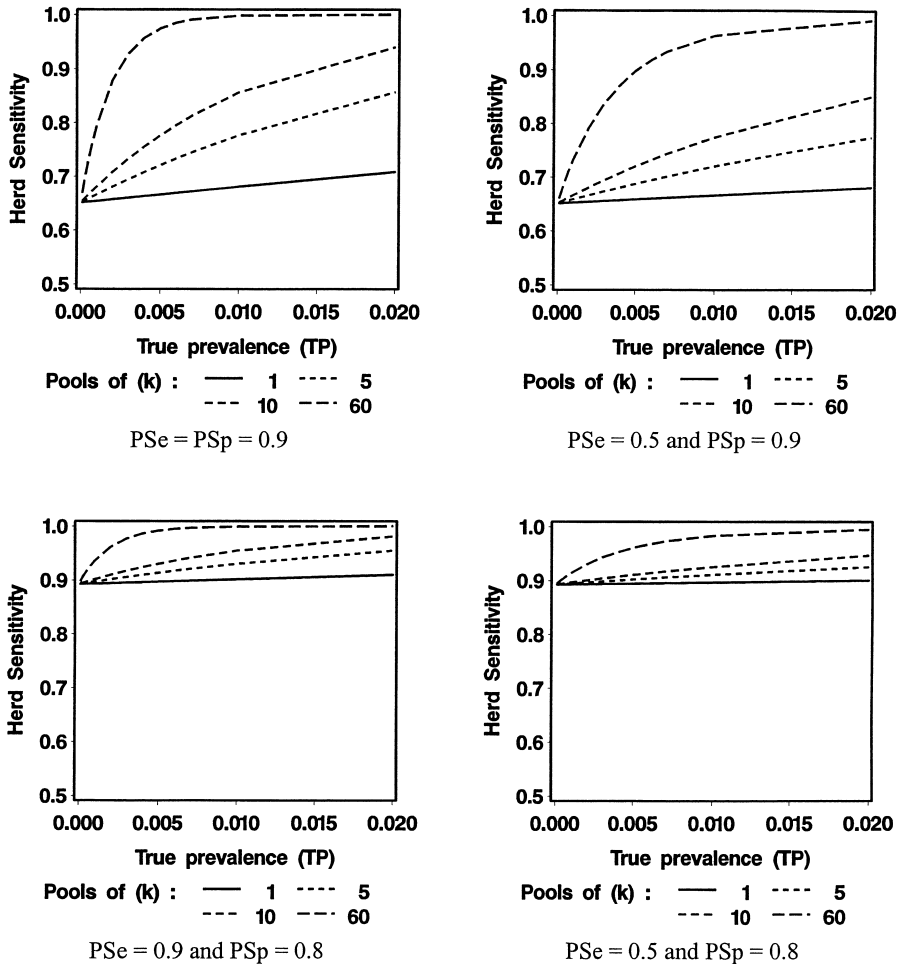


Fig. 8. The relationship between herd-level sensitivity estimates (for a herd-level test defined by 10 pools with the cut-off of 1 positive pool) and the true prevalence (TP) for different pool sizes (k) and pooled test characteristics (PSe and PSp).

For a herd test involving r pools from a non-infected herd, HPSp is dependent only on PSp and equals PSp^r . Hence, we have the same relationship between HPSp and PSp as we do between HSp and Sp.

When TP is low (<0.02), HPSe increases with increased number of animals included in the pools and decreased PSp (Fig. 8). However, the HPSp will decrease substantially with decreases in PSp. For PSp of 0.9 and 0.8 and $r=10$, the estimated HPSp (or PSp^r) are 0.35 and 0.11, respectively. The effect of PSe (decreasing from 0.9 to 0.5) on HSe is negligible for all k investigated (1–60).

Note that if $TP=0$ (and hence $HTP=0$), then herd apparent prevalence based on use of pooled tests ($HAPP$)= $1-HPSp=1-PSp^r=1-0.9^{10}=0.65$ (Fig. 8, upper curves) and 0.89

(Fig. 8, lower curves). Therefore, the unobserved PSp can be estimated from HAPP and equals $(1 - \text{HAPP})^{1/r}$, assuming no clustering of PSp. Furthermore, the new estimate of PSp can be used to estimate Sp (which is also unobserved). Assuming no clustering of Sp among pools, then $\text{Sp} = \text{PSp}^{1/k}$. For example, assume that a survey was done using $r=10$ pools per herd and $k=10$ animals per pool and one positive pool was sufficient to designate each herd as positive. In the survey, 65 of 100 herds were test positive with one positive pool per herd, i.e. $\text{HAPP}=0.65$ (Fig. 8). The PSp is estimated as $(1 - 0.65)^{1/10} = 0.35^{1/10} \approx 0.9$. Similarly, Sp can be estimated from PSp and equals $0.9^{1/10} \approx 0.99$.

3.5. Targeted or prevalence-directed sampling

Targeted or prevalence-directed sampling is another option to increase HSe when TP is expected to be low. This option is applicable if the only goal is to detect infection in a herd and if prior knowledge of clustering by identified factors, e.g. sex, age, building, breed or clinical signs can be used when sampling animals for the herd test. The additional information about the clusters is used to reduce the sampling frame from the entire herd to a proportion of the herd — thereby increasing the prevalence in the subpopulation that is sampled for the herd test. A benefit of targeted sampling is that a herd diagnosis can often be established with fewer samples. For example, in an outbreak investigation when samples are selected for culture from animals with typical lesions at necropsy (TP is close to 1), few samples will be necessary. In a control program to detect subclinical salmonella infection in a swine herd, the sampling frame could be reduced from the entire herd to the population of feeder and finisher pigs because TP (and hence HSe) is likely to be higher in pigs aged 3–9 months than in sows (Table 1).

3.6. Determining herd status with combinations of tests

Use of a combination of tests is another method to improve HSe or HSp, although incorporation of more tests will increase the cost of obtaining the desired HSe and HSp.

Table 1

A theoretical example of detection of subclinical salmonella infection in an infected Danish swine herd^a

Age-group ^b	Herd size	AP	Herd test performance	
			HSe = $1 - (1 - \text{AP})^b$	HSp = Sp^b
Sows	275	0.02	0.18	0.99
<20 kg	700	0.05	0.4	0.99
20–50 kg	750	0.10	0.65	0.99
>50 kg	775	0.15	0.8	0.99
Total	2500	0.09	0.62	0.99

^a The apparent prevalence (AP) is expected to differ by age-group; bacteriological examination for Salmonella is hypothesised to have Se=0.5 and Sp=0.999. Ten samples ($n=10$) are included in the herd test and the herd cut-off value=1. For simplicity, we assume that Se and Sp are unbiased and known for calculation of herd sensitivity (HSe) and specificity (HSp) values.

^b The distribution of pigs by age-group was based on the known distribution of age-groups in the Danish swine industry (Danske Slagterier, 1997).

When multiple tests are used for individual diagnosis, parallel interpretation (positive on at least one test is positive; negative otherwise) is used to increase Se and serial interpretation (positive on all tests is positive; negative otherwise) is used to increase Sp. For herd-level interpretation, a number of possible strategies could be used to define herd status when multiple tests are used. For example, if two binary tests are available for use on all animals, then two possible diagnostic approaches are to:

- interpret test results jointly for each individual (with either serial or parallel interpretation), then aggregate the results for all tested animals and interpret the results according to a decision threshold (herd cut-off value);
- define results of each herd test separately and then determine a decision rule for combining herd-test results (with either series or parallel interpretation).

The HSe and HSp will depend on the choice of diagnostic approach and the extent of conditional dependence between tests (see Gardner et al. (2000) for a description of conditional dependence at an individual-animal level) but a full extension of these issues to the herd level is beyond the scope of this paper. We demonstrate the concepts with two examples involving HSe and HSp and a fixed cut-off value of 1 positive-test result to designate the herd as positive. Other cut-offs would yield different conclusions (but, for brevity, we do not consider the effect of choice of cut-off value).

3.6.1. Herd specificity

Consider two tests (denoted as T_1 and T_2) both used on a sample of 10 truly negative animals and each yielding one positive test result — this result would be expected if both tests have $Sp=0.9$. Four pairs of test results are possible (T_1+T_2+ ; T_1+T_2- ; T_1-T_2+ ; T_1-T_2-) but two scenarios (A and B) are likely when the results for all animals are jointly considered. The likelihood of A or B is determined by the dependence between test specificities — scenario A is more likely when the Sp are conditionally independent and scenario B is more likely when the Sp are conditionally dependent.

A : $T_1 + T_2-$; $T_1 - T_2+$; $T_1 - T_2 -$ (1, 1, and 8 animals, respectively)

B : $T_1 + T_2+$; $T_1 - T_2 -$ (1 and 9 animals, respectively)

The first approach (interpret results jointly for each individual) in parallel would yield two and one false-positive test result for scenarios A and B, respectively. With serial interpretation, there would be zero and one false-positive test result, respectively. With a herd cut-off of 1, serial interpretation when tests are independent (scenario A) would yield a negative herd-test result and correctly classify the herd as negative (highest HSp). Serial interpretation with dependent tests (scenario B) or parallel interpretation regardless of test dependence (scenarios A and B) would misclassify herd status (false-positive herd test). The second approach (each herd test interpreted separately and then combined) would also misclassify herd status with either series or parallel interpretation because both would yield a false-positive result. Hence, in the second approach neither test dependence nor the method of interpretation affects the final conclusion about herd status.

3.6.2. Herd sensitivity

Now consider that both T_1 and T_2 are used on a sample of 10 animals from a truly infected herd and each test yields one positive result ($AP=0.1$ for both tests, but we do not specify the Se and Sp values). Again, there are two likely scenarios (A and B) for the combined test results, whose likelihood is determined by the conditional dependence between the test sensitivities and specificities. Scenario A is more likely than B when tests are independent and the opposite is more likely when the tests are dependent.

A : $T_1 + T_2-$; $T_1 - T_2+$; $T_1 - T_2 -$ (1, 1, and 8 animals, respectively)

B : $T_1 + T_2+$; $T_1 - T_2 -$ (1 and 9 animals, respectively)

The first approach (interpret results jointly for each individual) in parallel would result in two and one positive individuals (scenarios A and B, respectively) and a true-positive herd test. Interpreted in series, scenario A would falsely misclassify the herd as negative (lowest HSe) and scenario B would correctly classify herd status (a true-positive herd test). Again the second approach (each herd test interpreted separately and then results of the two herd tests are combined) would correctly classify the herd as positive. In contrast to the first example, all test interpretation methods and dependence structures — except for serial interpretation when tests are independent — would correctly classify the herd as positive.

4. Conclusions

We offer some concluding comments on the new aspects and considerations:

- Low precision of individual-animal Se and Sp estimates in evaluation studies will result in very wide (0–1) confidence intervals for HSe and HSp.
- Bias in individual-animal Se and Sp estimates can result in substantial bias in HSe and HSp estimates.
- Calculations of HSe and HSp done by currently available computer software assume that estimates of Se and Sp are unbiased and known with certainty. We recommend that investigators consider bias and sample sizes in validation studies of individual-test performance when considering the magnitude and certainty of calculated HSe and HSp values.
- Most risk-factor studies aim to make inferences about factors that are associated with disease. Therefore, we recommend use of the herd-test result rather than disease status defined by a control program in risk-factor studies, wherever possible.

Both pooled and targeted sampling can be cost-effective methods to improve herd sensitivity:

- The gain in HSe from use of pooled over individual-animal testing depends on factors such as true prevalence, pool size, number of pooled tests and the comparative performance of pooled and individual tests. Pooled testing is likely to have the maximum effect on HSe when true prevalence is low.

- Targeted sampling may increase HSe without decreasing HSp because HSp is unaffected by within-herd true prevalence. However, the method requires additional information on clustering of infection if it is to be practically applied.

Multiple tests may be combined to improve herd-test performance although costs of obtaining the desired HSe and HSp will be greater than when single tests are used. Gains will depend on whether serial or parallel interpretation is used and the conditional dependence between the tests. Complete conditional dependence in test sensitivities and specificities will not change HSe and HSp regardless of method of interpretation — and will only increase the cost of herd classification.

Acknowledgements

We thank Henrik Stryhn, Danish Veterinary Laboratory, for his help with and comments on the sections about precision and bias. We thank David Jordan and Angus Cameron for helpful comments about their computer programs (HERDACC and FREECALC, respectively).

Appendix A. Terminology and abbreviations

Test related

Individual test	individual-level test
Herd test	herd-level test
Individual cut-off	selected cut-off for an individual-test result
Herd cut-off	the number (or proportion) of positive individual tests selected for interpretation of the aggregate herd-level test as positive
T ₊ and T _−	test positive and test negative, respectively
HT ₊ and HT _−	herd test positive and negative, respectively
Se and Sp	individual-test sensitivity and specificity, respectively
HSe and HSp	herd-test sensitivity and specificity, respectively, based on individual-test results
PSe and PSp	pooled test sensitivity and specificity, respectively
HPSe and HPSp	herd-test sensitivity and specificity, respectively, based on pooled test results
HPV ₊ and HPV _−	predictive values of positive and negative herd-test results, respectively

Disease and prevalence

D ₊ and D _−	diseased and not diseased, respectively
TP	true within-herd prevalence based on individual-test results
TPP	true within-herd proportion of positive pools (at least one animal is D ₊)
HTP	proportion (prevalence) of truly positive herds
AP	apparent within-herd prevalence based on individual-test results

APP	apparent within-herd proportion of positive pools (at least one animal T+)
HAP	apparent proportion of positive herds based on individual-animal test results
HAPP	apparent proportion of positive herds based on pooled test results

Sampling

<i>n</i>	number of samples in a herd test
<i>r</i>	number of pools in a herd test
<i>k</i>	number of individual samples per pool
<i>N</i>	number of samples in evaluation study for Se
<i>M</i>	number of samples in evaluation study for Sp

Standard

Pr	probability
C	confidence
CI	confidence interval

References

- Acorda, J.A., Yamada, H., Ghamsari, S.M., 1995. Comparative evaluation of fatty infiltration of the liver in dairy cattle by using blood and serum analysis, ultrasonography, and digital analysis. *Vet. Quart.* 17, 12–14.
- Cameron, A.R., Baldock, F.C., 1998a. A new probability formula for surveys to substantiate freedom from disease. *Prev. Vet. Med.* 34, 1–17.
- Cameron, A.R., Baldock, F.C., 1998b. Two-stage sampling in surveys to substantiate freedom from disease. *Prev. Vet. Med.* 34, 19–30.
- Carpenter, T.E., Gardner, I.A., 1996. Simulation modeling to determine herd-level predictive values and sensitivity based on individual-animal test sensitivity and specificity and sample size. *Prev. Vet. Med.* 27, 57–66.
- Carrington-Reid, M., Lach, M.S., Feinstein, A.R., 1995. Use of methodological standards in diagnostic test research. Getting better but still not good. *J. Am. Med. Assoc.* 274, 645–651.
- Carstensen, B., Christensen, J., 1998. Herd size and sero-prevalence of *Salmonella enterica* in Danish swine herds: a random-effects model for register data. *Prev. Vet. Med.* 34, 191–203.
- Cowling, D.W., Gardner, I.A., Johnson, W.O., 1999. Comparison of methods for estimation of individual-level prevalence based on pooled samples. *Prev. Vet. Med.* 39, 211–225.
- Danske Slagterier, 1997. *Statistics 1996*. Danske Slagterier, Copenhagen, Denmark, p. 5.
- Donald, A.W., Gardner, I.A., Wiggins, A.D., 1994. Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors. *Prev. Vet. Med.* 19, 167–187.
- Gardner, I.A., Stryhn, H., Lind, P., Collins, M.T., 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev. Vet. Med.* 45, 107–122.
- Geishauser, T.D., Leslie, K.E., Duffield, T.F.B., Edge, V.L., 1998. An evaluation of protein/fat ratio in first DHI (Dairy Herd Improvement Corporation) test milk for prediction of subsequent displaced abomasum in dairy cows. *Can. J. Vet. Res.* 62, 144–147.
- Greiner, M., Gardner, I.A., 2000a. Application of diagnostic tests in veterinary epidemiologic studies. *Prev. Vet. Med.* 45, 43–59.
- Greiner, M., Gardner, I.A., 2000b. Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3–22.
- Griffin, J.M., Martin, S.W., Thorburn, M.A., Eves, J.A., Hammond, R.F., 1996. A case-control study on the association of selected risk factors with the occurrence of bovine tuberculosis in the Republic of Ireland. *Prev. Vet. Med.* 27, 217–229.

- Hoorfar, J., Bitsch, V., 1995. Evaluation of an O-antigen ELISA for screening cattle herds for *Salmonella typhimurium*. *Vet. Rec.* 137, 374–379.
- Jordan, D., 1996. Aggregate testing for the evaluation of Johne's disease herd status. *Aust. Vet. J.* 73, 16–19.
- Jordan, D., McEwen, S.A., 1998. Herd-level test performance based on uncertain estimates of individual test performance, individual true prevalence and herd true prevalence. *Prev. Vet. Med.* 36, 187–209.
- King, V.L., Koketsu, Y., Reeves, D., Xue, J., Dial, G.D., 1998. Management factors associated with swine breeding-herd productivity in the United States. *Prev. Vet. Med.* 35, 255–264.
- Losinger, W.C., Garber, L.P., Hill, G.W., Dornseif, S.E., Rodriguez, J.M., Frye, W.B., 1997. Design and implementation of the United States National Animal Health Monitoring System 1994–95 cattle on feed evaluation, and an evaluation of the impact of response biases. *Prev. Vet. Med.* 31, 1–14.
- Martin, S.W., Shoukri, M., Thorburn, M.A., 1992. Evaluating the health status of herds based on tests applied to individuals. *Prev. Vet. Med.* 14, 33–43.
- Mousing, J., Jensen, P.T., Halgaard, C., Bager, F., Feld, N., Nielsen, B., Nielsen, J.P., Bech-Nielsen, S., 1997. Nation-wide *Salmonella enterica* surveillance and control in Danish slaughter swine herds. *Prev. Vet. Med.* 29, 247–261.
- Nielen, M., Aarts, M.G.A., Jonkers, A.G.M., Wensing, T., Shukken, Y.H., 1994. Evaluation of two cowside tests for the detection of subclinical ketosis in dairy cows. *Can. Vet. J.* 35, 229–232.
- Ransohoff, D.F., Feinstein, A.R., 1978. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N. Eng. J. Med.* 299, 926–930.
- Rogan, W.J., Gladen, B., 1978. Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* 107, 71–76.
- Rogers, R.J., Cook, D.R., Ketterer, P.J., Baldock, F.C., Blackall, P.J., Stewart, R.W., 1989. An evaluation of three serological tests for antibody to *Brucella suis* in pigs. *Aust. Vet. J.* 66, 77–80.
- Smitsaart, E.N., Zanelli, M., Rivera, I., Fondevila, N., Compaired, D., Maradei, E., Bianchi, T., O'Donnell, V., Schudel, A.A., 1998. Assessment using ELISA of the herd immunity levels induced in cattle by foot and mouth disease oil vaccines. *Prev. Vet. Med.* 33, 283–296.
- Sørensen, V., Barfod, K., Feld, N.C., 1992. Evaluation of a monoclonal blocking ELISA and IHA for antibodies to *Mycoplasma hyopneumoniae* in SPF-pig herds. *Vet. Rec.* 130, 488–490.
- Sørensen, V., Barfod, K., Feld, N.C., Vraa-Andersen, L., 1993. Application of enzyme-linked immunosorbent assay for the surveillance of *Mycoplasma hyopneumoniae* infection in pigs. *Rev. Sci. Tech. Off. Int. Epiz.* 12, 593–604.
- Sørensen, V., Ahrens, P., Barfod, K., Feestra, A.A., Feld, N.C., Friis, N.F., Bille-Hansen, V., Jensen, N.E., Pedersen, M.W., 1997. *Mycoplasma hyopneumoniae* infection in pigs: duration of the disease and evaluation of four diagnostic assays. *Vet. Microbiol.* 54, 23–34.
- Vancini, V.R., Aguirre, N., Lugaresi, C.I., de Echaide, S.T., de Canavesio, V.G., Guglielmo, A.A., Marchesino, M.D., Nielsen, K., 1998. Evaluation of an indirect ELISA for the diagnosis of bovine brucellosis in milk and serum samples in dairy cattle in Argentina. *Prev. Vet. Med.* 36, 211–217.