Demonstrating disease freedom—combining confidence levels

R.M. Cannon*

Office of the Chief Veterinary Officer, Agriculture, Fisheries and Forestry—Australia, P.O. Box 858, Canberra, ACT 2601, Australia

Received 3 January 2001; accepted 25 October 2001

Abstract

Part of the requirements for demonstrating disease freedom usually will be that sufficient testing be done to give a specified confidence of detecting the disease if it were present at a specified level. Often, this requirement is translated into a fixed testing regime that must be followed (an inflexible approach that might not be the most economic or practical solution).

A more flexible approach is to specify the capabilities of the various tests that can be used to detect the disease, and let the party hoping to demonstrate disease freedom decide upon the testing regime. The question then arises as to how to combine information that can come from a variety of sources over a period of time to give an overall level of confidence.

Two methods are given. The first, an exact method based on multiplying probabilities, would be more appropriate for a survey of an area in which no disease is thought to be present. The second method (more appropriate for a herd-assurance program within an infected area) is a point-based system that takes into account the different sensitivities of the methods used to detect disease and the change in prevalence over time. It allocates points for each test done proportional to the sensitivity of the test and the prevalence at the time of testing. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Sampling; Confidence levels; Disease freedom; Assurance program; Voluntary accreditation program

1. Introduction

1.1. Overview

The need to demonstrate ‘disease freedom’ occurs at various levels, from an individual herd to an entire country. However, many regulations specifying the requirements for
demonstrating disease freedom are inflexible because they specify the testing requirements (type of test to be used, number of herds to be tested and number of animals to be tested within herd). Such codes concentrate on ‘how to do it’ rather than on ‘what needs to be shown’.

The characteristics of the disease would be reflected in the specifications for demonstrating disease freedom. These might include: the minimum prevalence (at both herd and within-herd levels) that any testing is to detect, the confidence required of doing so, a minimum period for the surveillance, a minimum interval until a herd can be re-tested, an adequate geographic spread of tested herds, and measures to be taken to guard against the introduction of new infection. Accepted minimum sensitivities and standard techniques for known tests could be included in the requirements as an addendum with new tests easily added.

The important thing is that the choice of tests and the location and amount of the sampling does not need to be specified. The herd’s owner or a country’s administration should be able to determine which methods of testing would be used after considering the cost of a testing regime, the market requirements, the need to demonstrate the absence of disease quickly, and other economic and practical factors.

As an example (used in Section 5), consider how to demonstrate that a herd is free of Johne’s disease. The disease has a long incubation period and is difficult to detect, especially in its early stages. There are a number of different tests available (with different costs, different (low) sensitivities, and different problems associated with false positives). Some tests might be done as part of a market-assurance program that regularly tests the herd. Other information might come from inspecting animals at slaughter. Some testing might be done in response to clinical observation. Indeed, the fact that clinical observation has detected no disease contributes in its own small way to the confidence that the herd might be free. In addition, the prevalence of the (undetected) disease would be expected to increase over time, and so become easier to detect. Some farmers might want to reach a ‘free’ status (or a lesser status) as quickly as possible. Others might prefer to take longer by using cheaper, less-intrusive methods. How can all these different items be combined into a simple system to demonstrate herd freedom—and, even further, that a zone (a region) is free of disease?

Consider sample size, prevalence and confidence—specify two of these for a simple survey and the third is determined. Section 2 discusses some considerations required when specifying these and other factors related to survey design. The point emphasised is that ‘what needs to be shown’ should be separated from ‘how to show it’ to provide a flexible, economic and practical survey regime that can combine the information from a number of sources.

Equations for combining confidence levels are given in Section 3. In general, these are mathematically simple (often, not more than multiplying the error probabilities associated with the tests done). They can be computed easily in a survey of a zone, in which there is a single target for the entire zone. The situation is different when considering a number of individual herds—each trying to reach a target to achieve freedom. Many people are less comfortable with a target based on the multiplication of small fractions than they are with a need to reach a certain number of points. The point-based system (described in Section 4) was devised as an approximation; the emphasis is on simplicity at the user level, at the
expense of statistical rigour. The point system assigns a fraction of a point for each test done (the value being proportional to the sensitivity of the test and the likely prevalence at the time of the test). Points are subtracted for events that reduce the confidence that the herd is disease-free. If no infected animals are detected, the total number of points accumulated by a herd from all its testing provides a measure of the confidence level that the herd is free of disease. Criteria that define what a herd can do would be defined in terms of points achieved. Section 5 completes the paper with an example of how a point scheme could be used.

2. Definitions, assumptions and notation

2.1. Round of testing

The paper considers confidences at two levels—herd and zone. It will be convenient to have a lower level corresponding to a ‘round’ of testing. A round of testing is an occasion on which a number of animals in a herd are tested at about the same time by the same method. The number of animals tested in a round could range from a single animal to the whole herd.

2.2. Confidence and sensitivity

The terms ‘confidence’ and ‘sensitivity’ are effectively synonymous. Both describe the probability that a disease would be detected if it were present, although the context of each is usually different: sensitivity refers to components (e.g. tests) and confidence refers to the overall result (e.g. round). However, in going from animal to round to herd to zone, one level’s confidence becomes the sensitivity used to calculate the next level’s confidence. For example, when calculating the zone-level confidence, the herd-level sensitivity used is equal to the herd-level confidence achieved by the rounds of testing.

The symbol TSe will be used to denote a test’s sensitivity, with RSe, HSe, and ZSe being used for particular levels of sensitivity—round, herd, and zone—or confidence as the context requires. The symbol γ will be used to denote confidence when a general use of the term is required.

2.3. False positives

Serological tests can give occasional false-positive reactions. In general, the design of any survey to demonstrate freedom should specify a sequence of further testing that would be done to clarify the true status when a positive reaction is detected and questioned. Such a sequence would effectively result in a 100%-specific test.

Sometimes it is not possible to do other tests to confirm positive tests because, for example, the cost would be prohibitive or no further material is available (as might occur when taking blood samples at an abattoir). In such cases, it is necessary to specify a tolerance limit for the number of positive results that will differentiate between the false positives in a non-infected herd and the combinations of true and false positives in an infected herd. For example, Cameron and Ballock (1998) and Cannon (2001) give
methods that relate the number of tests and tolerance to the round-level sensitivity and specificity. In terms of the number of tests done, it is best if a round of testing has an extremely high specificity.

In developing the point system, this paper assumes that any positive tests are resolved in such a way that there will be no false-positive rounds or herds.

2.4. An appropriate level of confidence

The requirements for demonstrating disease freedom may include a serological survey that is designed, typically, with a 95% confidence level of detecting disease if it were there. Setting such a confidence level implicitly takes into account the unquantified confidence provided by many other factors such as clinical observation, naive population or conservative (low) estimates of disease prevalence. Increasingly, there are attempts to incorporate into the overall confidence many of these factors, especially when the prevalence would be very small. Once a value has been placed on this extra confidence, the overall confidence level required as proof should be much higher than 95%.

2.5. Design prevalence

The confidence that a round of testing would detect disease depends on the prevalence at the time of testing. This, of course, is not known. A value called the ‘design prevalence’ needs to be specified and used (both to determine the sample size required to detect the disease with the desired confidence and to determine the level of confidence that results from a particular sample size that was not based on a formal sample-size calculation). The precise meaning of the design prevalence in a long-term survey can be difficult to define—for example, it might be a constant minimum (expressed as a number or percentage) or a parameter of a binomial or some other distribution—because it is dynamic (depending on sales, deaths, births, purchases and the disease’s spread). The value for the design prevalence can be chosen in several ways depending on the purpose of the survey, but (like any parameter) will be chosen conservatively, in a way that would result in a larger sample size and hence increase the chance of detection.

For some uses (such as a market-assurance or an accreditation program), it may be sufficient to assign arbitrarily a low value for the design prevalence (say 2%) although even that figure might be chosen as a compromise between the confidence given and the cost of the testing required to achieve it. Such a program would do simply what it was designed to do: provide the desired confidence that a herd with the design prevalence would be detected. Herds with a prevalence greater than the design prevalence would be detected with more confidence; those with a prevalence lower might still be detected, but with less confidence.

In other uses (such as demonstrating freedom of disease in a zone), the design prevalence would be based on the level of disease likely to be present in a herd and a zone if it had been present—but undetected—for sometime. A value for the design prevalence could be based on a conservative (low) estimate from field observations in endemic zones or a low percentile from disease modelling. For such uses, the design prevalence would still be conservative, but it need not be small. For example, if a highly contagious disease like
porcine reproductive and respiratory syndrome had spread through a herd unnoticed, more than 20% of the pigs could be expected to have antibodies.

The herd size indirectly affects how the design prevalence is expressed. For large herds, the prevalence can be set as a proportion. However, for small herds with a low prevalence, it is often better to specify the prevalence as a number of diseased animals (because it is not possible to have a fraction of a diseased animal).

The symbols $P$ and $HP$ will be used for within-herd prevalence (proportion of animals infected in a herd) and herd-level prevalence (proportion of herds infected in a zone), respectively. The symbols $DP$ and $DHP$ will be used for the design (within herd) prevalence and for the design herd-level prevalence, respectively.

2.6. Number of animals to test

When demonstrating a zone is free from infection, there is considerable flexibility in setting the number of herds tested and number of tests per herd that will meet the requirements of the survey. This flexibility means that other criteria (such as convenience or cost) can be used to decide on sample numbers (see, for example Cameron and Baldock, 1998 or Cannon, 2001). Determining the best method to use would include considering the location of testing (e.g. farm, abattoir or a combination of both) and the financial implications of the test or tests to be used (e.g. amount of sensitivity per dollar and average extra cost to resolve possible false positives). What might be an optimal solution in one jurisdiction need not be the best in another.

3. Combining confidence levels

This section gives formulas for calculating the overall confidence resulting from a number of methods of testing at the round, herd and zone levels. The probability that two independent rounds of testing both fail to detect a disease is equal to the product of the probabilities that each round would fail to detect the disease by itself. Hence, the overall confidence $\gamma$ provided by two independent rounds with confidence $\gamma_1$ and $\gamma_2$ is given by

$$\gamma = 1 - [(1 - \gamma_1)(1 - \gamma_2)].$$

This basic formula is used repeatedly (at the test-, round- and herd-levels) to combine confidences. Occasionally, it is used in reverse to determine the amount of extra testing needed to be done. If an overall confidence $\gamma$ is required, and the first round of testing gave a confidence of $\gamma_1$, the confidence $\gamma_2$ required for the second round is

$$\gamma_2 = 1 - \frac{1 - \gamma}{1 - \gamma_1} = \frac{\gamma - \gamma_1}{1 - \gamma_1}.$$ 

3.1. Multiple tests on the same animal

At its lowest level, Eq. (3.1) could be used to determine the overall sensitivity when multiple tests are done on a single animal. However, often Eq. (3.1) is not applicable
because the test results on the same animal are correlated (typically because the tests measure the same response to the disease). Generally, experimental data would be needed to estimate the combined sensitivity (or specificity). This (together with binomial confidence limits) can be obtained from the proportion of infected (or non-infected) animals correctly categorised by the rules for combining the test results (e.g. both positive or either positive), a more direct approach than Gardner et al. (2000, Sections 3.1 and 3.2).\(^1\)

Several authors use the terminology ‘conditional dependence of tests’ to describe the relationship between tests. I think ‘correlated tests’ is better because it simply implies a degree of agreement between the tests, caused by some factor, rather than a dependency of one test on the other test. The use of the word ‘conditional’ is both misleading and unnecessary. It links more naturally to the word ‘dependence’ rather than the intended meaning ‘dependence of tests conditional on the infection status of the animal’. However, the definitions of sensitivity and specificity already imply the infection status of the animals tested. The correlation between the results of testing infected animals might be totally different to that when testing non-infected animals.

### 3.2. Round-level confidence

The level of confidence provided by a round of testing (RSe) depends on the number of animals tested (\(n\)), the test-level sensitivity (TSe), and the within-herd prevalence (\(P\)). The form of the relationship depends on how the prevalence is interpreted.

If the herd size is large, the apparent prevalence equals the test sensitivity multiplied by the prevalence (\(P \times \text{TSe}\)). The round-level confidence is given by the probability that testing detects no infected animals (among \(n\) animals tested):

\[
\text{RSe} = 1 - (1 - P \times \text{TSe})^n.
\]

(3.3)

If the finite size of the population needs to be taken into account (with \(P = D/N\), the proportion of diseased animals (\(D\)) in the herd of \(N\) animals), then an approximation (Cannon, 2001) for the round-level confidence is

\[
\text{RSe} \approx 1 - \left(1 - \frac{n \times \text{TSe}}{N - \frac{1}{2}(D \times \text{TSe} - 1)}\right)^D.
\]

(3.4)

The values of TSe and \(P\) used to calculate the round-level confidence would be those applicable at the time the round of testing was done, and might be different from round to round. Modelling could provide one way of estimating how the prevalence might vary with time, see, for example Schlosser and Ebel (2001).

If a round of testing consists of \(n\) pooled tests in which samples from \(k\) animals are combined and tested as one, a formula similar to Eq. (3.3) is used but depends on the definition of pool-level sensitivity (PSe). For example, in discussing pooled

\[^1\text{When discussing the agreement between two tests, Gardner et al. (2000) consider both the ratio of observed covariance to maximum possible covariance and kappa, a common measure of agreement. A third measure, the correlation (\(\rho\)), seems more natural because (using the notation of their 2 \times 2 contingency Table 3), \(\rho^2 = (ad - bc)^2/\{(a + b)(c + d)(a + c)(b + d)\}, \ (a + b + c + d) \rho^2 = \chi^2\text{ statistic used for testing independence. Their formula for covariance simplifies to } \gamma_{\text{Se}} = (ad - bc)/(a + b + c + d).\} \]
tests, Christensen and Gardner (2000) define PSe as the probability of detecting infection in a pool in which at least one animal is infected. However, with this definition, the sensitivity will vary with the number of infected animals in the pool, which in turn will depend on the within-herd prevalence. It is important that a value for PSe is always qualified by a prevalence—either the number of infected animals and pool sizes used when evaluating the test or the assumed within-herd prevalence when the test is being used.

When considered as a herd test, it may be better to define PSe as the probability that infection is detected in a pool if the herd (rather than the pool) is infected. In a large herd, the difference between the two definitions is a factor of $1 - (1 - P)^k$, corresponding to the loss of sensitivity because no infected animals contribute to the pool. With this definition, the round-level sensitivity is derived using Eq. (3.1) as

$$RSe = 1 - (1 - PSe)^n.$$  

(3.5)

PSe, based on herd prevalence (HP), often can be calculated from the experimental values by combining the probabilities of each possible number of infected animals in a pool and the corresponding probability of detection. For example, if a test can detect one infected animal in a pool of $k$ animals with probability Se, and two or more infected animals in the pool with probability 1, the PSe of the test can be calculated as

$$PSe = 1 - (1 - P)^k - (1 - Se) \times k \times P \times (1 - P)^{k-1}$$

(3.6)

by considering the probability that 0 or 1 infected animals contribute to the pool.

### 3.3. Herd-level confidence

At a herd level, the confidence gained from a number of independent rounds of testing (HSe) can be calculated from the individual round-level sensitivities RSe as

$$HSe = 1 - \prod (1 - RSe_i).$$

(3.7)

The product in this formula is taken over all sources of information for that herd, such as an allowance for competent clinical observation, any miscellaneous testing, herd tests or abattoir monitoring. Any variation in within-herd prevalence over the period of testing is taken into account in the calculation of RSe.

For some diseases and tests, Eq. (3.7) can overestimate the confidence because of the correlation between successive rounds of tests when the same animals are re-tested. This might arise if the response measured by the test takes a long time to develop in the animal or if an infected animal consistently gives a false-negative result. If this is thought to be an important problem, it would be necessary to combine the confidence based on all testing for each animal (past or present) in the herd (ASe) to calculate HSe:

$$HSe = 1 - \prod (1 - P \times ASe_i).$$

(3.8)

### 3.4. Zone-level confidence

The confidence that a zone (with $N$ herds) is free of disease is the probability that the testing would have detected disease in the zone (ZSe). It depends on
• the amount of testing done on each herd in the zone and the design within-herd prevalence (which are reflected in the herd-level sensitivities, HSe); and
• the DHP or number of diseased herds \( D = DHP \times N \)—the minimum level of disease that would be expected if the disease indeed were present.

The exact formula for the zone-level confidence can be derived by a simple combinatorial argument. There are \( N \times (N - 1) \times \cdots \times (N - D + 1) \) possible ways that the \( D \) infected herds can be chosen from the \( N \) herds in the zone. For each possible way, the probability that the disease would have been detected can be calculated. The sum of all these probabilities gives the zone-level sensitivity:

\[
ZSe = 1 - \sum_{i_1=1}^{N} \sum_{i_2 \neq i_1} \cdots \sum_{i_D \neq i_1, i_2, \ldots} \frac{(1 - HSe_{i_1}) \times (1 - HSe_{i_2}) \times \cdots \times (1 - HSe_{i_D})}{N \times (N - 1) \times \cdots \times (N - D + 1)}.
\]

(3.9)

The above equation is complicated to use. However, by considering the average herd-level sensitivity \( HSe_{ave} \) for the zone, there is a simple approximation (equivalent to the binomial approximation of the hypergeometric distribution). The probability of detecting the first diseased herd is \( HSe_{ave} \). The probability of detecting the second diseased herd will be approximately \( HSe_{ave} \)—the actual value is the average herd-level sensitivity of all but the first diseased herd—and so on for the third, fourth, etc. diseased herds. An adequate approximation for the probability of detecting at least one of the \( D \) diseased herds is

\[
ZSe \approx 1 - (1 - HSe_{ave})^D.
\]

(3.10)

An alternative approach is to interpret the DHP as the parameter of a binomial distribution-equivalent to treating the number of herds in the zone as infinite. If the disease were present, each herd would have the same probability DHP of being infected, and the probability that the disease is detected in a particular herd will be DHP multiplied by the herd-level sensitivity. The overall confidence would be

\[
ZSe = 1 - \prod_{i=1}^{N} (1 - DHP \times HSe_i).
\]

(3.11)

The product is taken over all herds in the zone. Any herd that had no testing will have \( HSe_i = 0 \), and so contributes nothing to the confidence.

If an allowance for the increase of HP over time were to be made, it would be necessary to calculate the herd-level sensitivities and zone-level confidence on, say, an annual basis, and then combine the annual confidences using Eq. (3.1).

For some types of testing, the herd of origin might not be routinely available. Even so, the testing still contributes to the zone-level confidence. If only a few animals per herd are tested, the approximate confidence \( \gamma \) gained from testing \( n \) animals is found by assuming that each animal has the same probability \( (DHP \times DP \times TSe) \) of reacting to the test if the disease were present:

\[
\gamma \approx 1 - (1 - DHP \times DP \times TSe)^n.
\]

(3.12)

When a large number of animals per herd is tested, the calculation of the approximate confidence gained needs to be done in two steps, both using Eq. (3.4). The first step
calculates the herd-level confidence based on the estimated average number of animals per herd tested, and the second uses this value with the estimated number of herds tested.

4. A point system for combining confidence

The point-based system described in this section is an approximate way of combining the confidence from a number of different types of tests. It is designed for people who regularly need to know the overall level of confidence by providing a simpler way of measuring confidence and uses a more intuitive target than the exact methods of the previous section. Typically, a point system could be the basis of an assurance (or accreditation) program by herds wishing to sell animals, or could be used to confirm that an infected herd thought to have eradicated the disease had indeed done so. This introduction gives a broad description of such a system, and leaves the precise details for gaining points to Sections 4.1–4.3 and for losing points to Sections 4.5–4.10. An example of how such a system might work is given in Section 5.

No known-infected herd would be in the system. A herd that came in contact with any animals or material from a known-infected herd would leave the system (at least temporarily, until its status was resolved).

For the simplest implementation, each round of testing contributes points equal to the number of tests multiplied by the sensitivity (as a fraction) of the test. Provided that no infected animals are detected, the total number of points—the herd’s score—gained from all testing provides a measure of the confidence level that the herd is free. If the disease is regulated, target scores would determine what the herd could or could not do. However, decisions about the type and timing of tests would be left to the herd owner to make. A point-based system is well suited to a de-regulated environment. A purchaser could make decisions based on the vendor’s score (a continuum of values rather than two or three defined status levels in typical schemes). The system also can adjust the points gained to take into account changes in within-herd prevalence over time.

If the point system is being used within an infected area, the possibility of disease entering the herd should be considered. Because there is always a small chance that a herd for which there is no evidence of infection is actually infected, purchasing animals reduces the confidence that a herd is free, and so its score should be decreased. This is especially so if a herd acquires its animals from a large number of herds. The number of points lost depends on the number of animals involved and the difference between the scores of the two herds involved. Because of the logarithmic relationship between confidence and points, calculating the number of points lost does not involve simple arithmetic, but is easily done in a spreadsheet. Although the designers of the system will have to deal with the more complicated formulas that follow, the users will simply see the system as a set of tables giving the points gained for each type of test or points lost as the consequence of purchasing animals.

4.1. Gaining points

The point system gives a fraction of a point for each animal tested. The fraction is equal to the sensitivity of the test multiplied by a weighting equal to the prevalence at the time of
the test divided by the DP. As a starting point to this informal justification of the system, consider the number of animals that need to be tested by a perfect test to be able to detect a specified prevalence of disease (the DP) with the desired confidence. How should this number be changed to take into account lack of sensitivity and changes in prevalence over time?

Because a 50% sensitive test only detects disease half as often as a 100% sensitive test, it seems reasonable to assume that about twice as many animals should be tested with the poorer test than with the perfect test. An equivalent approach (which is the basis of the point system) is to keep the target the same, but only count half an animal for each 50% sensitive test done. By keeping a constant target and valuing each tested animal as a fraction of an animal equal to the sensitivity of the test, the problem of combining the confidence provided by tests with different sensitivities is overcome.

In the first few years after the introduction of a slowly spreading disease like Johne’s disease, the prevalence is likely to be less than the DP. If it is appropriate to take into account changes in prevalence as a disease spreads through a herd or zone, the same approach can be used—halving the prevalence halves the chance of detecting the disease and doubles the number of tests required. Again, this is equivalent to keeping the target constant and valuing each test as a fraction of an animal (namely, the prevalence at the time of testing expressed as a fraction of the DP).

These approximations form the basis of the point system. The number of points \(S_i\) assigned for the \(i\)th round of testing takes into account the number of tests \(n_i\), a value for the test \(v_i\) related to its sensitivity, and the prevalence at the time of the test \(P_i\). It is convenient to express \(P_i = DP \times w_i\) in terms of a DP and a weighting factor \(w_i\). The round of testing would contribute \(S_i\) points to the total herd score of \(S\) points:

\[
S = \sum S_i, \quad \text{where } S_i = n_i \times v_i \times w_i.
\]

(4.1)

For a round in which single animals were being tested, \(v_i\) would be equal to the sensitivity of the test, and \(n_i\) would be the number of animals tested. If the round of testing were a whole-herd test, \(v_i\) would be determined from the herd-level sensitivity of the test (see Section 4.3) and \(n_i\) would be 1. For a pooled test in which a number of animals are tested simultaneously, \(v_i\) would be derived from the PSe, and \(n_i\) would be the number of pools tested.

The weighting factors \(w_i\) can reflect changes in the within-herd prevalence of a herd over time or it can be used to allow for differences in DP between herds. It may be difficult to specify values for weighting factors. Use \(w_i = 1\) if weighting factors are not applicable.

For simplicity, the point system does not take into account any correlation between successive tests on the same animal although a minimum inter-test interval should be specified. If testing routinely consists of a pair of tests (e.g. a screening test and a definitive test), the points gained would be based on the combined sensitivity of the two tests.

4.2. Relationship between score, prevalence and confidence

The point system uses the following approximation between the score \(S\), the DP and the confidence \((HSe)\) that disease would be detected if the disease was present:

\[
HSe = 1 - \exp(-DP \times S), \quad \text{(or equivalently)} \quad S = -\frac{\ln(1 - HSe)}{DP}.
\]

(4.2)
To derive this result, first combine Eqs. (3.7) and (3.3) (using \(v_i\) for the test sensitivity, and \(w_i \times DP\) for the prevalence at the time of the test) to give the herd-level confidence in terms of the amount of testing

\[
1 - HSe = \prod (1 - RSe_i) = \prod (1 - w_i \times DP \times v_i)^{n_i},
\]

and then take (natural) logarithms to convert from multiplications to additions:

\[
\ln(1 - HSe) = \sum n_i \times \ln(1 - w_i \times DP \times v_i).
\]

Because \(w_i \times DP \times v_i\) is small, and \(\ln(1 - x) \approx -x\) for small \(x\), the approximate relationship between \(HSe\) and \(S\) is

\[
\ln(1 - HSe) \approx -\sum n_i \times w_i \times DP \times v_i = -DP \times S,
\]

which is another way of writing Eq. (4.2).

The herd’s score can be used to estimate an upper \(\gamma\)% confidence limit to the proportion of infected animals \((P_{\text{upper}})\) in the herd. Replace \(HSe\) with \(\gamma\) and \(DP\) with \(P_{\text{upper}}\) in Eq. (4.2) to give

\[
P_{\text{upper}} = -\frac{\ln(1 - \gamma)}{S}.
\]

Alternatively, a target might be set as the number of points \((T)\) that must be achieved to demonstrate freedom at the desired confidence \((\gamma)\). Because \(T\) and \(\gamma\) are related by Eq. (4.2)

\[
T = -\frac{\ln(1 - \gamma)}{DP}.
\]

For a given value of \(DP\), this target is the same regardless of herd size, which is unrealistic for small herds. A target proportional to the herd size would be more appropriate for small herds (less than DN animals, where DN would need to be chosen) and this can be achieved by taking the design within-herd prevalence as a fixed number DN \(\times DP\) of infected animals. Because it is more convenient to have the same target for each herd, a weighting factor of DN\(N\) would be used in a small herd to adjust the score for the greater prevalence.

Table 1 illustrates Eqs. (4.6) and (4.7). For example, with 300 points the upper 95% confidence limit for the within-herd prevalence is 1%. To achieve a confidence level of 99%

\begin{table}
\centering
\begin{tabular}{lcccccc}
\hline
\textbf{Prevalence} & \multicolumn{4}{c}{\textbf{Confidence}} & \multicolumn{2}{c}{\textbf{Score}} \\
& 1% & 2% & 5% & 100% & 100 & 200 & 300 & 400 & 500 \\
\hline
1% & 230.3 & 115.1 & 46.1 & 2.3026 & 90.0 & 2.30 & 1.15 & 0.77 & 0.58 & 0.46 \\
2% & 299.6 & 149.8 & 59.9 & 2.9957 & 95.0 & 3.00 & 1.50 & 1.00 & 0.75 & 0.60 \\
5% & 460.5 & 230.3 & 92.1 & 4.6052 & 99.0 & 4.61 & 2.30 & 1.54 & 1.15 & 0.92 \\
10% & 496.2 & 248.1 & 99.2 & 4.9618 & 99.3 & 4.96 & 2.48 & 1.65 & 1.24 & 0.99 \\
1% & 690.8 & 345.4 & 138.2 & 6.9078 & 99.9 & 6.91 & 3.45 & 2.30 & 1.73 & 1.38 \\
\hline
\end{tabular}
\caption{The relationship between score, prevalence and confidence. The score \((S)\), prevalence \((P)\) and confidence \((\gamma)\) are related by \(P \times S = -\ln(1 - \gamma)\).}
\end{table}
using a DP of 1%, we need to set a target of 460.5 points. For simplicity, we might increase this target to 500 (equivalent to 99.3% confidence) and use 500 as the dividing point between large and small herds. For small herds, a weighting factor of \( w_i = \frac{500}{N} \) for points scored would be used, based on \( 5 = 500 \times 1\% \) infected animals in the herd.

4.3. Converting the confidence provided by a pooled test into points

Some tests (for example, a bulk milk test) are whole-herd tests that have a herd-level sensitivity rather than an individual animal-level sensitivity. The sensitivity \( (TSe_p) \) of the test, which might vary with the prevalence \( (P) \) at the time of testing, can be converted to a value \( (v_i) \) using Eq. (4.2):

\[
v_i = -\frac{\ln(1 - TSe_p)}{P}.
\]  

(4.8)

For example, a whole-herd test with a sensitivity of 40% when the within-herd prevalence is 2% would be worth 25 points \( (25.24 = -\ln(0.6)/0.02) \). Before being added to the score, these points would be multiplied by the weighting \( (w_i) \) and by \( n_i = 1 \) (for the number of tests). The same formula can be used to value pooled tests in which samples from several animals are combined before testing. The number of pooled tests would be used as the multiplier in calculating the score. The values for each type of test would be tabulated.

4.4. A zone-based point system

The point system can be used on a zone basis, although the exact methods of Section 3.4 would seem more appropriate. It is important to note that the herd scores cannot simply be added together to give a score for the zone. A zone’s score is the weighted sum of the annual herd-level sensitivities \( (HSe_{i,y}) \) that result from the amount of testing in herd \( i \) during year \( y \), reflected in each individual herd’s score \( (S_{i,y}) \):

\[
S = \sum_{years} w_y \sum_{herds} HSe_{i,y}, \quad \text{where } HSe_{i,y} = 1 - \exp(-DP \times S_{i,y}).
\]  

(4.9)

The weighting factors \( (w_y) \) take into account changes in the HP over the period of surveillance.

It is also possible to collect zone points from testing in which it is not practical to collect herd identification routinely—although of course the herds themselves would not gain points. Abattoir monitoring (with follow-up property identification of any diseased animals detected) would be a typical example. Allocation of zone points would reflect the variability of herd-submission patterns and the level of statistics being kept. If an average of \( A \) animals from a herd are tested whenever a herd is tested, the round-level sensitivity would be \( 1 - (1 - TSe \times DP)^A \) from Eq. (3.3). This many zone points would be gained for each herd tested, with the number of herds tested either being known or estimated from the total number of animals tested. Strictly speaking, a slightly lesser number of points should
be added if a herd is tested a second time, but the slight difference would be ignored for simplicity.

4.5. Losing points

If the point system (or any assurance program) is used in a known-infected area, the possibility of disease entering the herd must be considered. The risk could be a specific event such as the purchase of apparently healthy animals or something non-specific such as a background risk per year. In either case, the confidence that the herd is not infected would be reduced, and the score should be decreased.

To some extent, a loss of points is contrary to the expectation and purpose of many assurance schemes in which herds of a similar status can trade without testing and without degrading their status. Nonetheless, purchasing animals provides a chance of introducing infection (even though the risk is small if the purchased animals come from a herd with a high score). Purchasing animals will reduce the confidence that the herd is disease-free and should result in the loss of points. The introduction of or contact with animals from a known-infected herd would take the herd outside the point system until its status is resolved.

For herds outside the point system, losing points because of purchases is irrelevant—the point system has served its purpose by providing a measure of the risk associated with the purchased animals. The loss of points from purchasing animals is really only relevant to a herd that routinely sells animals. Such a herd would be expected to make greater efforts to ensure that disease does not enter the herd, and purchasers should expect the points to reflect any risks taken. Testing the purchased animals would generate points that could be added to the vendor’s score but not to the purchaser’s score. However, the purchaser would benefit from any testing by way of a lesser risk from the purchase and hence a lesser reduction of points.

The next four sections provide a justification for the rules for losing points. These rules are based on what happens to the scores of two herds when they merge. The concept is extended to any challenge by using the relationship between a herd’s score and the probability that the herd is infected to express the risk as an equivalent number of points. These points will be called the ‘merging value’ of the challenge: the greater the merging value, the greater the confidence that there is no risk, and the less the reduction to the herd’s points. The system has been devised so that the number of points lost depends only on the difference between the herd’s score and the merging value of the risk. If the merging value of a risk is less than the herd’s score, the score drops to the merging value, and then further points are lost based on the difference between the two values. Unfortunately, while the relationship between the difference and number of points lost is simple, to write down, its calculation requires more than simple arithmetic and would be implemented instead as a table or spreadsheet.

4.6. Points and probability that the herd is infected

Once points can be lost, it becomes more meaningful to treat the score as a measure of the probability that the herd is not infected, rather than the confidence of detecting disease. Without any testing, the probability the herd is infected is equal to the HP of the zone.
After testing, the probability ($\delta$) that a herd is infected can be found (by Bayes rule) from the resultant herd-level confidence ($\text{HSe}$), based on the DP and the herd’s score ($S$):

$$\delta = \text{HP} \times \frac{1 - \text{HSe}}{1 - \text{HSe} \times \text{HP}}. \quad (4.10)$$

By substituting $1 - \text{HSe} = \exp(-\text{DP} \times S)$ from Eq. (4.2), it is possible to determine the value of $S$ that corresponds to a probability $\delta$. However, to simplify the point system, the probability that the herd is infected is approximated as being equal to the probability that a random herd is infected ($\text{HP}$) multiplied by the probability that the infection has not been detected:

$$\delta \cong \text{HP} \times \exp(-\text{DP} \times S). \quad (4.11)$$

By taking logarithms, the probability $\delta$ can be expressed as an equivalent number of points $M$:

$$M = \frac{-\ln(\delta/\text{HP})}{\text{DP}}. \quad (4.12)$$

$M$ will be positive if $\delta < \text{HP}$, 0 if $\delta = \text{HP}$ and negative if $\delta > \text{HP}$. The point system uses Eqs. (4.11) and (4.12) to switch between probabilities (risks) and herd scores. The number of points ($M$) will be called “the merging value of the risk”—a name that is a little incongruous for something like a background risk, but more realistic when the risk comes from purchased animals being merged into the herd.

Many risks will be proportional to HP, so a precise value for HP is not needed. For example, if no neighbouring herds are known to be infected, the probability of being infected by a neighbour would be proportional to HP, the zone prevalence (and proportional to the number of neighbours). As an example of Eq. (4.12), if DP = 0.01, a risk of $\delta = 0.02 \times \text{HP}$ would have a merging value of 391.2 = $-\ln(0.02)/0.01$. However, there are risks that are not proportional to HP, such as the risk associated with a known-infected neighbouring herd. There is a problem for a free zone in which HP = 0, because Eq. (4.12) is undefined. A pragmatic approach needs to be taken, and a small value assigned to HP for that purpose. In practice, the results of Eq. (4.12) would be tabulated simply as a list of events and the corresponding merging values.

### 4.7. Merging two herds

Suppose that two herds with $S_1$ and $S_2$ points are merged and that $S_1 \leq S_2$. To determine the score $S$ of the combined herd, the point system ignores the very small probability that both herds are infected and approximates the probability that the combined herd is infected as the sum of the small probabilities that each herd is infected. By using Eq. (4.11) to convert scores to probabilities this gives

$$\text{HP} \times \exp(-\text{DP} \times S) \cong \text{HP} \times \exp(-\text{DP} \times S_1) + \text{HP} \times \exp(-\text{DP} \times S_2). \quad (4.13)$$

Taking logarithms gives a simple formula for $S$ that is suitable for direct calculation

$$S = -\frac{\ln(\exp(-\text{DP} \times S_1) + \exp(-\text{DP} \times S_2))}{\text{DP}}. \quad (4.14)$$
However, a result that is more easily tabulated is obtained by multiplying both sides of Eq. (4.13) by \( \exp(\frac{DP}{C^2}S_1) \) before taking logarithms. The combined score \( S \) is the minimum of the two scores \( (S_1) \) minus a function \( \text{LostPts} \) of the difference between the higher and lower score \( (S_2 - S_1) \):

\[
S = S_1 - \text{LostPts}(S_2 - S_1), \quad \text{where} \quad \text{LostPts}(X) = \frac{\ln[1 + \exp(-DP \times X)]}{DP}.
\] (4.15)

Fig. 1 illustrates the relationship between the difference of the scores and the number of points lost, although in practice the result might be tabulated rather than graphed. For example, if \( DP = 1\% \) (bold) and \( DP = 2\% \) (italic). Example: if \( DP = 1\% \), and herds with scores of 300 and 450 merge, the difference is 150. From the graph, the loss would be \(-20\) giving a score of 280 for the combined herd.

4.8. The risk of purchasing infected animals

Consider the purchase of \( n \) animals, each of which is tested by a test with sensitivity \( T_{Se} \) (taken to be zero if no test is done). If any infected animals were detected, all \( n \) animals would be rejected. The merging value associated with purchasing these animals is more
than the herd’s score because there is a chance—which decreases as \( n \) increases—that there will be no infected animals in the group even though the source herd is infected.

Before calculating the risk of this purchase, the vendor’s score would be increased to \( S \) to include points gained from the testing associated with the sale. The purchaser has effectively purchased \( U = n \times (1 - TSe) \) animals that have not been tested, although if the vendor’s score is low, \( U \) needs to be adjusted as described in Section 4.10. The risk is the product of three probabilities: the probability the herd is infected (HP), the chance that infection would not have been detected if the vendor’s score is \( S \), and the probability of undetected infected animals being in the ‘\( U \) untested animals’:

\[
\delta \cong HP \times \exp(-DP \times S) \times (1 - \exp(-DP \times U)).
\]  

(4.16)

This may be converted using Eq. (4.12) to give the merging value \( M \) equal to the herd’s score plus some ‘part herd points’ (PHP) that are a function of the number of untested animals purchased:

\[
M = S + \text{PHP}(U), \quad \text{where} \quad \text{PHP}(U) = \frac{-\ln(1 - \exp(-DP \times U))}{DP}.
\]  

(4.17)

Fig. 2 illustrates the relationship between the number of untested animals and the PHP. The number of PHP decreases to zero as \( U \) increases. \( M \) would be infinite if \( U = 0 \), resulting in no loss of points.

Table 2 gives examples of the number of points lost from purchasing animals. The results are not unexpected: the greater the number of animals purchased, the greater the loss of points; also, the lower the vendor’s score, the greater the loss of points. The points lost are small for a herd with few points, but can be quite considerable for a purchaser with a high score. Whether animals should be purchased from a herd with a significantly lower score is an economic decision to be made by the herd owner rather than being something prescribed by regulators.

---

Fig. 2. The part herd points \( (V) \) associated with purchasing \( U \) untested animals. The function graphed is \( V = -\ln(1 - \exp(-DP \times U))/DP \) for a \( DP = 1\% \) (bold) and \( 2\% \) (italic). The part herd points are used to calculate the merging value \( (M) \) associated with the purchase of \( n \) animals all of which test negative to a test with sensitivity \( Se \): (1) calculate the vendor’s new score \( S \) as a result of the tests, (2) calculate the number of untested animals:

- \( DP = 1\% \) and \( S < 100 \), \( U = n \times (1 - Se) \times 100/S \),
- \( DP = 2\% \) and \( S < 50 \), \( U = n \times (1 - Se) \times 50/S \),
- otherwise, \( U = n \times (1 - Se) \).

(3) look up \( U \) on the graph and read off \( V \) from the opposite side of the graph, and (4) calculate \( M = S + V \). Example: if \( DP = 1\% \) and \( U = 10 \) untested animals are purchased from a herd with score \( S = 200 \), look up \( V = 236 \) (opposite 10), and add \( S \) and \( V \) to give a merging value \( M = 436 \).
If no infected animals are detected, any testing of the vendor’s herd after the sale of animals increases the confidence that the purchased animals were free of disease. It would be possible to recalculate \( M \) (and reduce the points lost because of a purchase) to take into account later testing of the vendor’s herd.

### 4.9. Adjusting \( U \) for vendors with low scores

An adjustment to the value of \( U \) used in Eq. (4.17) should be made if the vendor’s score is small \((S < \text{DP})\) because the maximum chance of purchasing infected animals occurs if the prevalence is more than DP. Consider the purchase in terms of the point system as \( S \) animals tested with a ‘100% sensitive test’ and \( U \) untested animals. Using Eq. (4.2), the probability \( \eta \) that there were no infected animals in the \( S \) tested animals but some infected animals in the \( U \) untested animals will vary with the actual, but unknown, within-herd prevalence \((P)\):

\[
\eta \approx \exp(-P \times S) \times (1 - \exp(-P \times U)).
\]

This probability is zero for \( P = 0 \), rises to a maximum and then decreases to zero. Using calculus, the maximum occurs when

\[
P = \frac{\ln(U + S) - \ln(S)}{U} \approx \frac{1}{S}.
\]

For small \( S \), this maximum occurs for a value of \( P \) greater than DP and so, in the term \( \exp(-\text{DP} \times U) \) in Eq. (4.16), DP should be replaced by \( 1/S \) (the maximum value of \( P \)) to give \( \exp(-U/S) \). This may be written as \( \exp(-\text{DP}/\text{DP} \times U/S) \), which is equivalent to using Eq. (4.17) with \( U \) adjusted (increased):

\[
U_{\text{adj}} = \begin{cases} 
\infty & \text{if } S < 0, \\
\frac{U}{\text{DP} \times S} & \text{if } S < \left(\frac{1}{\text{DP}}\right).
\end{cases}
\]

### Table 2

The reduction in score after purchasing animals

<table>
<thead>
<tr>
<th>(a) Score after purchasing the equivalent of 10 untested animals</th>
<th>Purchaser’s score</th>
<th>(b) Score after purchasing untested animals from a herd with 300 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor’s score</td>
<td>Number of untested animals purchased</td>
<td>5</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>−12.6</td>
<td>177.0</td>
<td>190.9</td>
</tr>
<tr>
<td>−4.8</td>
<td>246.8</td>
<td>277.0</td>
</tr>
<tr>
<td>−1.7</td>
<td>293.1</td>
<td>346.8</td>
</tr>
<tr>
<td>−0.6</td>
<td>317.6</td>
<td>393.1</td>
</tr>
<tr>
<td>0</td>
<td>335.2</td>
<td>435.2</td>
</tr>
</tbody>
</table>

\( a \) The table shows how the vendor’s score and the number of untested animals purchased affect the purchaser’s score. The left-hand side of the table keeps the number of animals purchased constant and varies the vendor’s score, while the right-hand side does the reverse. A DP of 1% was used.
For example, if $U = 6$, $S = 25$ and $DP = 2\%$, the adjustment of Eq. (4.20) would be used because $S < 50$ to give $U_{\text{adj}} = 12$.

4.10. Comparing scores in different zones

Herds with the same score from zones with different HPs will have the same confidence that infection would have been detected, but will have different absolute probabilities of being infected. Consequently, before using Eq. (4.11), the point system must adjust the merging value calculated by Eq. (4.17) when moving animals from zones of different prevalences. The necessary adjustment (based on the merging value $M_V$ and prevalence HP$_V$ in the vendor’s zone) can be derived by rewritten Eq. (4.12) as

$$M = \frac{-\ln((\delta/\text{HP}_V)(\text{HP}_V/\text{HP}))}{\text{DP}} = M_V \frac{-\ln(\text{HP}_V/\text{HP})}{\text{DP}}. \quad (4.21)$$

For example, with $DP = 2\%$, when a herd in a zone with $\text{HP} = 1\%$ purchases animals from a zone with $\text{HP} = 5\%$, the merging value needs to be reduced by $80.4 = \ln(0.05/0.01)/0.02$ points to be comparable. In practice, a rounded value of 80 would be used.

An alternative approach is to build the difference defined by Eq. (4.21) into the system by giving a positive initial score for herds in lower-prevalence zones. Based on the example above, herds in a zone with $\text{HP} = 5\%$ could start with 0 points, but herds in a zone with $\text{HP} = 1\%$ would start with 80 points. There would be no need to adjust for differences between zones. However, the adjustment for untested animals in Eq. (4.20) must be changed to reflect the non-zero starting point $S_0$ in the vendor’s zone:

$$U_{\text{adj}} = \begin{cases} \infty & \text{if } S < S_0, \\ \frac{U}{\text{DP} \times (S - S_0)} & \text{if } S < S_0 + \left(\frac{1}{\text{DP}}\right). \end{cases} \quad (4.22)$$

5. Example and comments

Discussion of the point system will centre on the following example of an ovine Johne’s disease-control program. The program’s aim is to reduce the spread of the disease between properties with a minimum of regulations and restrictions on the sale of animals. The point system would be used to quantify the testing. The example is an idealised scheme that is not in operation anywhere. Its purpose is to demonstrate how the point system might work rather than to provide an answer to each and every question relating to its implementation. In particular, the values used for the sensitivity of the various tests and other tabulated items are purely indicative and would be set according to local conditions. The basic rules of such a system might be:

- until a flock has been inspected by an accredited veterinarian, no points can be scored;
- the score is increased whenever confidence is increased (e.g. negative test results);
- the score is reduced whenever confidence is reduced (e.g. purchases);
- a flock that is accumulating points can purchase sheep only from a flock with a score;
• an infected flock is outside the point system until the disease is thought to be eradicated and has restrictions on the sale (but not the purchase) of animals; and
• a separate (but virtually identical) scheme would apply to infected flocks that have undergone an eradication program or flocks that must be classified as infected because of contact with infected animals but are thought not to be infected because of the slight degree of contact.

Table 3 is a hypothetical example illustrating how points could accumulate in a flock over a 4-year period. The first column of Table 3 gives the date of the event, as a way to identify lines in this discussion. Column 2 gives the type of entry for the line, either a test type or the reason for loss of points (with the merging value in column 3). Columns 4–6 give the number of tests, the point value for a single test, and the weighting to take into account changes in prevalence. The product of these—the total number of points corresponding to the round of testing—is in column 7. Column 8 accumulates the points gained over time, and the scores have been converted to a confidence (column 9) using the approximation of Eq. (4.2) with a DP of 1%.

5.1. Initial inspection

Before a flock can gain any points, a veterinarian must visit the property to obtain a history of the flock and to rule out any obvious infection (at least with a visual inspection of all animals, and possibly by testing some animals).

The general purchasing pattern would determine the initial score. A flock which has introduced animals from a large number of sources—particularly from areas of high prevalence—could be considered to have a higher risk of being infected than the background prevalence for the area, and could be assigned a negative score appropriate to the risk.

The history of the flock would determine whether there should be a weighting penalty to allow for low within-flock prevalence. If there were no reason to suppose that the within-flock prevalence would be different from the DP for a typical undiagnosed infected flock in the area, the weighting would be 1 from the start. It might be prudent to discount any points accumulated during the first year (e.g. by 50%) as protection against dishonesty.

However, if a flock had a known ‘last date of possible infection’, it is more realistic to assume that the prevalence would start from zero at that date. If it were thought that the DP might be reached in 3 years, a possible set of weightings could be 0.1 during the first year, 0.3 during the second year, 0.6 during the third year and 1.0 thereafter (as in column 6 in this example).

The initial inspection would confirm that boundary security was adequate to provide some protection against the introduction of animals (and possible infection) from neighbouring properties. This would be used to determine the risk because of possible accidental introduction of disease (see Section 5.3).

5.2. Accumulating points

The sensitivities underlying the values in Table 3 are for illustrative purposes only. For example (September 2001), 300 ELISA, with a sensitivity of 50% would only count for
Table 3
Illustration of points accumulated by a herd over a 4-year period of testing*

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Merging value</th>
<th>n</th>
<th>v</th>
<th>w</th>
<th>n × v × w</th>
<th>Score</th>
<th>Confidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 2000</td>
<td>Initial visit</td>
<td>–</td>
<td>1</td>
<td>10.00</td>
<td>0.1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.00</td>
</tr>
<tr>
<td>January 2000</td>
<td>Post-mortem</td>
<td>–</td>
<td>2</td>
<td>1.00</td>
<td>0.1</td>
<td>0.2</td>
<td>1.2</td>
<td>1.19</td>
</tr>
<tr>
<td>January 2000</td>
<td>Pooled faecal culture</td>
<td>–</td>
<td>2</td>
<td>28.00</td>
<td>0.1</td>
<td>5.6</td>
<td>6.8</td>
<td>6.57</td>
</tr>
<tr>
<td>May 2000</td>
<td>Abattoir monitoring (1 in 5)</td>
<td>–</td>
<td>300</td>
<td>0.06</td>
<td>0.1</td>
<td>1.8</td>
<td>8.6</td>
<td>8.24</td>
</tr>
<tr>
<td>January 2001</td>
<td>Background risk</td>
<td>460</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–1.1</td>
</tr>
<tr>
<td>January 2001</td>
<td>Visual inspection</td>
<td>–</td>
<td>1</td>
<td>10.00</td>
<td>0.3</td>
<td>3.0</td>
<td>10.5</td>
<td>9.98</td>
</tr>
<tr>
<td>June 2001</td>
<td>Pooled faecal culture</td>
<td>–</td>
<td>7</td>
<td>28.00</td>
<td>0.3</td>
<td>58.8</td>
<td>69.3</td>
<td>50.00</td>
</tr>
<tr>
<td>September 2001</td>
<td>ELISA</td>
<td>–</td>
<td>300</td>
<td>0.50</td>
<td>0.3</td>
<td>45.0</td>
<td>114.3</td>
<td>68.12</td>
</tr>
<tr>
<td>September 2001</td>
<td>Post-mortem</td>
<td>–</td>
<td>5</td>
<td>1.00</td>
<td>0.3</td>
<td>1.5</td>
<td>115.8</td>
<td>68.59</td>
</tr>
<tr>
<td>October 2001</td>
<td>Purchase animals</td>
<td>602</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–0.8</td>
</tr>
<tr>
<td>November 2001</td>
<td>Abattoir monitoring (1 in 3)</td>
<td>–</td>
<td>250</td>
<td>0.10</td>
<td>0.3</td>
<td>7.5</td>
<td>122.5</td>
<td>70.64</td>
</tr>
<tr>
<td>January 2002</td>
<td>Background risk</td>
<td>460</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–3.4</td>
</tr>
<tr>
<td>January 2002</td>
<td>Visual inspection</td>
<td>–</td>
<td>1</td>
<td>10.00</td>
<td>0.6</td>
<td>6.0</td>
<td>125.2</td>
<td>71.40</td>
</tr>
<tr>
<td>January 2002</td>
<td>Pooled faecal culture</td>
<td>–</td>
<td>7</td>
<td>28.00</td>
<td>0.6</td>
<td>117.6</td>
<td>242.8</td>
<td>91.18</td>
</tr>
<tr>
<td>July 2002</td>
<td>Abattoir monitoring (1 in 3)</td>
<td>–</td>
<td>250</td>
<td>0.10</td>
<td>0.6</td>
<td>15.0</td>
<td>257.8</td>
<td>92.41</td>
</tr>
<tr>
<td>January 2003</td>
<td>Background risk</td>
<td>460</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–12.4</td>
</tr>
<tr>
<td>January 2003</td>
<td>Visual inspection</td>
<td>–</td>
<td>1</td>
<td>10.00</td>
<td>1.0</td>
<td>10.0</td>
<td>255.3</td>
<td>92.22</td>
</tr>
<tr>
<td>March 2003</td>
<td>AGID</td>
<td>–</td>
<td>200</td>
<td>0.40</td>
<td>1.0</td>
<td>80.0</td>
<td>335.3</td>
<td>96.50</td>
</tr>
<tr>
<td>August 2003</td>
<td>Abattoir monitoring (1 in 5)</td>
<td>–</td>
<td>400</td>
<td>0.06</td>
<td>1.0</td>
<td>24.0</td>
<td>359.3</td>
<td>97.25</td>
</tr>
<tr>
<td>January 2004</td>
<td>Background risk</td>
<td>460</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–31.2</td>
</tr>
</tbody>
</table>

* The column numbers and dates are referred to in the text. The points gained for each test (in column 7) is the product of the number of tests (n), the value of the test (v) and the prevalence weighting factor (w). Events that lose points are given in italics with the points lost (in column 7) depending on the merging value of the risk and the herd’s previous score. The table is based on a DP of 1%.
45 points because the weighting for prevalence was only 0.3 at the time. A similar number of animals tested in the 4th year would generate 150 points. In the September 2001 testing, five positive ELISA reactors were found to be negative on post-mortem inspection and would gain extra points. Strictly speaking, the five reactors should not gain points for the ELISA.

Abattoir monitoring (in the form of inspection in the gastrointestinal tract for lesions) might have a sensitivity of 30% at the animal level. Because of the speed of the processing line, there would be only time to inspect a proportion of the animals. Hence, the sensitivity of an inspection effectively becomes 10% if every third animal is inspected, and 6% if only one in five is inspected (compare July 2002 with August 2003). The number of animals slaughtered (rather than the number of animals actually inspected) would be entered in column 4, with the value of the test reflecting the inspection rate.

A pooled faecal culture test (January 2002) using a single faecal pellet from each of 50 sheep might have a test sensitivity of 25% when the within-flock prevalence is 1%. This would correspond to 28 points \((= -\ln(1 - 0.25)/0.01)\) from Eq. (4.8).

A visual inspection of the flock by a veterinarian might score 10 points—partly as a token gesture, but also to reflect the fact that clinical inspection should detect chronic disease. In the example, in January 2000, points were scored for the initial inspection and the subsequent post-mortem inspection of two poor-condition animals.

5.3. Loss of points—maintaining status

The initial visit would assess practices in the flock that would affect the likelihood of disease entering the flock, e.g. animals entering the property from neighbours or routine movement of the flock on common land. If none of the neighbouring flocks were known to be infected, this risk would be proportional to (and expressed as a fraction of) the flock prevalence in the zone. The risk might be increased each year unless an accredited veterinarian reconfirms the adequacy of the perimeter security and other measures required to reduce the introduction of infection. If a neighbouring property is known to be infected, the risk would be much greater and would no longer depend on the flock prevalence in the zone. It could be that some risks are so great that the flock could not enter the scheme. It could be difficult to estimate a probability associated with some risks.

In this example, a risk equal to 1% times the flock prevalence has been used as a background risk, which is equivalent to a merging value of 460 from Eq. (4.12). The number of points lost each year (on the anniversary of entering the scheme) from the background risk becomes greater as the score increases because of the logarithmic relationship between score and confidence. Indeed, the merging value of the annual risk effectively places a limit to the target that a flock can reach and maintain.

5.4. Purchasing animals

A minor motivation for requiring that a flock in the scheme can purchase animals only from a flock with points would be to encourage flocks to join the system. A more pertinent reason is that a flock purchasing animals from a flock with zero points would lose all of its points (and more). The merging value (October 2001) from purchasing 10 animals tested
with a 50% sensitive test from a flock with 300 points is 602.1 \( (= 300 + 302.1 \) from Eq. (4.17).

A computerised system that keeps track of purchases between participating flocks would have benefits in tracing some of the sales from a flock subsequently found to be infected. It would also be possible to adjust (reduce) the points lost by a purchaser to take into account any later testing of the vendor’s flock.

5.5. Infected and suspect flocks

A flock that is infected or suspect would not be eligible to join the scheme until the disease is no longer thought to be present. In the case of a transient contact with infected animals, a small amount of testing might be sufficient to suggest that it is unlikely that the disease became established—although for a disease with a long incubation period this might take sometime. However, for an infected flock, it would be raising false hopes to accumulate points until an accepted eradication program had been followed.

As an example, a previously infected flock could be considered free (with all restrictions lifted) if it reaches a target of 500 points (which corresponds to a 99.3% confidence of detecting a prevalence of 1%).

Effectively, there would be two nearly identical systems: one would be an assurance scheme for flocks for which there is no reason to suppose that the flock is infected, and the other would confirm that an infected or suspect flock was no longer infected. There would be little difference between the two schemes, although there would certainly be different movement restrictions. The points for each type of test would be the same—but the weighting for prevalence could be different.

5.6. Other diseases

There are many factors that would determine if a point system might be useful for managing a particular disease. Although the epidemiology of the disease and the capability and number of available tests would be prime considerations, the urgency and the level of compulsion for demonstrating freedom would possibly be even more important. The point system would seem to be a good approach for a voluntary assurance program, but less so for a compulsory eradication campaign.

One would imagine a point system for bovine tuberculosis would be very similar to that just described, because like Johne’s disease, tuberculosis has relatively insensitive tests (e.g. skin test, gamma-interferon, inspection for granulomas at slaughter) that might only detect infection sometime after the animal was infected.

Contrast this with a disease like bovine brucellosis, which could be expected to spread quickly through a herd, and which can be detected soon after infection with cheap, highly sensitive blood tests. In an eradication campaign, herds that had shown no signs of the disease at an initial herd test might be required to have a confirmatory herd test after, say, 3 years unless sufficient testing had been done during the period. Such testing would be a combination of blood tests taken at abattoirs as part of regular monitoring, blood tests done for miscellaneous reasons such as sale of animals, and bulk milk-ring tests. Because the level of confidence from the testing only needs to be calculated once (at the end of the
period), an exact calculation would be more appropriate than a point system. Even if there were incentives to purchase animals from well-tested herds (which requires up-to-date knowledge of the amount of testing), it might be simpler (and sufficient) to require the testing of all animals being sold, rather than setting up a point system (because of the highly sensitive tests).

One of the strengths of the point system is its flexibility to combine the confidence provided by different types of tests, and possibly it would not be as useful if there was only one type of test available. However, the fact a herd loses points for purchases is an attractive feature because it reinforces to the herd owner the need to be aware of where animals come from without being too proscriptive—the herd owner must weigh the consequences of any actions undertaken.

6. Summary

The paper has given two methods for combining the confidence from various sources to give an overall confidence that there is disease freedom. The first method (using exact formulas) is more appropriate to a zone-based demonstration of disease freedom when there is not the need to know routinely the individual-herd contributions. The second (which accumulates point based on the amount of testing) is an approximation designed as a herd-based system in which individual herds can demonstrate their status in an infected area. The emphasis of both methods is the flexibility to choose the amount and type of testing according to their needs by those wishing to demonstrate disease freedom.

References