A new probability formula for surveys to substantiate freedom from disease

Angus R. Cameron a, F. Chris Baldock b,*

a Lao-Australian Animal Health Project, PO Box 7042, Vientiane, Laos
b AusVet Animal Health Services, 12 Thalia Court, Corindna, Queensland 4075, Australia

Accepted 12 September 1997

Abstract

Surveys to substantiate freedom from disease are becoming increasingly important. This is due to the changes in rules governing international trade in animals and animal products, and to an increase in disease eradication and herd-level accreditation schemes. To provide the necessary assurances, these surveys must have a sound theoretical basis. Until now, most surveys have been based on the assumption that the screening test used was perfect (sensitivity and specificity both equal to one), and/or that the study population was infinite. Clearly, these assumptions are virtually always invalid. This paper presents a new formula that calculates the exact probability of detecting diseased animals, and considers both imperfect tests and finite population size. This formula is computationally inconvenient, and an approximation that is simpler to calculate is also presented. The use of these formulae for sample-size calculation and analysis of survey results is discussed. A computer program, ‘FreeCalc’, implementing the formulae is presented along with examples of sample size calculation for two different scenarios. These formulae and computer program enable the accurate calculation of survey sample-size requirements, and the precise

Abbreviations: p, Prevalence; Se, Sensitivity; Sp, Specificity; D+, Disease-positive animals (true positives); D-, Disease-negative animals (true negatives); T+, Test-positive animals (positive reactors); T-, Test-negative animals (negative reactors); PO, Probability of an event with the event of interest described in the brackets; x, Number of T+ in a sample; y, Number of D+ in a sample; n, Sample size; N, Population size; d, Number of diseased D+ animals in the population; Bin(n, p), Binomial distribution with parameters n and p: \( \binom{n}{x} \), Number of ways that x objects can be drawn from n, equal to \( (n!)/(x!(n-x)!)) \)

* Corresponding author. Tel.: +617 3379 5385; fax: +617 3278 1953; e-mail: ausvet@eis.net.au

Disease in this context is defined in its broadest context: possessing the abnormality or state of interest. In surrogate tests for disease, it may mean, for example, the presence of antibodies.

0167-5877/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.

PII S0167-5877(97)00081-0
analysis of survey results. As a result, survey costs can be minimised, and survey results will reliably provide the required level of proof. © 1998 Elsevier Science B.V.

Keywords: Surveys; Freedom from disease; Aggregate testing; Detection of disease; Hypergeometric distribution

1. Introduction

Recent changes in regulations governing the international trade in animals and animal products have led to a greater need for epidemiologically sound surveys to substantiate the freedom of a country or zone from particular diseases. The World Trade Organisation (WTO) has the responsibility of implementing the various international agreements finalised through the Uruguay round of the General Agreement on Tariffs and Trade (GATT) (Anon., 1994). The WTO began operations in January 1995 and has adopted the codes of the Office International des Epizooties (OIE) to serve as guidelines for international trade in animals. A critical component of these guidelines is the establishment of regional, national, or sub-national disease-free zones for the purpose of livestock exports. Many countries are endeavouring to eradicate high-impact or trade-limiting diseases such as rinderpest, tuberculosis, foot and mouth disease (FMD), and contagious bovine pleuropneumonia. Substantiation of the final success of these campaigns will depend partly on statistically valid surveys. In addition, countries currently free from disease may also be required to provide stronger evidence for their disease-free status than simply the absence of clinical reports. Absolute proof requires the examination of every animal in the population, using an infallible test. As this is not feasible, sample surveys are used, and results are reported in terms of probability statements that the disease, if present, has a prevalence lower than a defined level that is usually determined by prior experience or analogy from other countries.

It is important to draw a distinction between surveys that aim to substantiate freedom from disease (or detect disease), and those that are designed to provide an estimate of disease prevalence. When surveying to substantiate freedom, the conclusion of the survey will be that disease is or is not present, and this statement will be qualified by the probability that it is correct. Surveys to estimate disease prevalence will produce an estimate of the prevalence, usually qualified by a confidence interval. The statistical theories behind the design and analysis of the two types of survey are different. Clearly, data gathered during the course of a survey to substantiate freedom from disease can be used to estimate the prevalence of disease, although differing survey design requirements may mean that the confidence intervals of the prevalence estimate are too wide for the intended use.

In the past, two assumptions have been commonly made to simplify the calculation of these probabilities, namely: (1) that the diagnostic test being used is perfect, and/or (2) that the population being studied is infinite (or that sampling is performed with replacement). Formulae based on these assumptions are presented below. However, virtually no diagnostic test is perfect, and populations under study are often small enough to be considered finite, especially when sampling animals within a herd. One other assumption of all common formulae is that sampling of the populations takes place
by simple random sampling. Violations of this assumption invalidate the results of the formulae. This important limitation is sometimes neglected—especially in the analysis of slaughterhouse surveys (McLaughlin, 1995), or surveys based on passively acquired serum banks (Thornton and Motha, 1995).

The sample size calculation that is commonly used estimates the number of animals that need to be tested (when sampling by simple random sampling from a homogenously mixing population), for the probability of selecting at least one diseased animal to be greater than a preset confidence level (commonly 95%). The calculation is based on the assumption that, if the disease is present, it will be present at a certain minimum prevalence.

The appeal of the approximate formula of Cannon and Roe (1982) for sample size is its simplicity. The results have been tabulated and published, and can be calculated for any population size using a pocket calculator. A simple modification of the approximate formula considering test sensitivity has been produced (MacDiarmid, 1988). The exact probability formula for finite populations and imperfect tests presented here is much more complex, and difficult to use. Calculation of the binomial approximation to this formula is somewhat faster, but still awkward. A major drawback is that the formula for probability cannot be equated to \( n \), the sample size. To calculate the sample size that produces a given probability, a process of trial and error is required. This process has been implemented in a computer program (‘FreeCalc’), which calculates the exact sample size, and can also be used as a probability calculator for the analysis of survey results.

This paper reviews the approaches in common use, and introduces two new formulae that can be used to plan and analyse surveys to substantiate freedom from disease. The use of a new computer program to simplify this task is described, and two examples of the practical application of the formulae for planning surveys are presented.

2. Probability formulae

2.1. Perfect test, infinite population

The probability of selecting a given number of reactors when selecting animals from a population with disease prevalence \( p \) is given by the binomial distribution:

\[
P(T^+ = x) = \binom{n}{x} p^x (1 - p)^{n-x}.
\]

If a perfect test is used, a survey to substantiate freedom from disease requires that no diseased animals are found. When \( x = 0 \), this formula simplifies to:

\[
P(T^+ = 0) = (1 - p)^n.
\]

2.2. Imperfect test, infinite population

When an imperfect test is used, the situation is more complex. The probability of getting a positive test result \( P(T^+) \) when testing a single animal depends on the true
disease status of that animal. If it is disease positive, \( P(T^+) \) is equal to \( Se \); if it is disease negative, \( P(T^+) \) is equal to \( 1 - Sp \). The overall probabilities \( P(T^+) \) and \( P(T^-) \) are given by:

\[
P(T^+) = pSe + (1-p)(1-Sp) \\
P(T^-) = p(1-Se) + (1-p)Sp.
\] (3)

The probability of observing \( x \) reactors when testing \( n \) animals from an infinite population is given by the binomial distribution (Eq. (1)) with the positive and negative probabilities substituted from Eq. (3):

\[
P(T^+ = x) = \binom{n}{x} [pSe + (1-p)(1-Sp)]^x [p(1-Se) + (1-p)Sp]^{n-x}.
\] (4)

2.3. Perfect test, finite population

In the case of finite population sizes, trials are not independent. When the first animal is drawn, the probability of drawing a \( D^+ \) is \( d/N \), but with each \( D^+ \) drawn, \( d \) is decreased by one. \( N \) decreases with every animal drawn. When a perfect test is used, the probability that \( T^+ \) (and \( D^+ \)) will equal \( x \) is given by the hypergeometric distribution:

\[
P(T^+ = x) = \frac{\binom{d}{x} \binom{N-d}{n-x}}{\binom{N}{n}}.
\] (5)

When aiming to substantiate freedom from disease, \( x \) is equal to zero, and this formula simplifies to (Cannon and Roe, 1982):

\[
P(T^+ = 0) = \frac{(N-d)!(N-n)!}{(N-d-n)!N!}.
\] (6)

Unfortunately, factorial formulae involving large numbers are awkward to calculate. Hand-held calculators can usually calculate factorials no higher than about 70!, and personal computers use logarithmic approximations to exceed factorials of about 170!. A convenient approximation to this formula, (e.g. Cannon and Roe, 1982) is:

\[
P(T^+ = 0) = \left(1 - \frac{d}{N - \frac{n-1}{2}}\right)^n.
\] (7)

2.4. Imperfect test, finite population

To overcome the limitations of the above formulae, the hypergeometric distribution (Eq. (5) above) can be modified for imperfect tests. The number of \( D^+ \) in the sample
has a hypergeometric distribution. Given a number \( y \) of \( D^+ \) in the sample, the number of true positives is \( \text{Bin}(y, Se) \), and the number of false positives is \( \text{Bin}(n - y, 1 - Sp) \). We will have a number \( x \) of \( T^+ \) if we have \( j \) true positives and \( x - j \) false positives. By considering the possible values of \( y \) and \( j \), we can write:

\[
P(T^+ = x) = \sum_{y=0}^{d} \binom{d}{y} \binom{N-d}{n-y} \sum_{j=0}^{\min(x,y)} \binom{y}{j} Se^j (1 - Se)^{y-j} \binom{n-y}{x-j} \times (1 - Sp)^{x-j} Sp^{n-x-y+j}.
\]

2.5. Approximation to the hypergeometric

2.5.1. Description of the approximation

Although able to provide the exact probability, Eq. (8) above is too complex to be practically used except for relatively small sample sizes. In addition to the problems of calculating large factorial terms, it is a summation that approaches \( 2^n \) terms. Even when implemented on a computer, calculation can be very slow. To overcome this difficulty, an approximation (described in Appendix A) was developed based on the binomial distribution. The approximation uses an estimate of the average probability of selection of a \( D^+ \) or \( D^- \) animal. This estimate is based on the assumption that the number of \( D^+ \) animals in the population, \( d \), is decreased by a non-integer amount at each selection, proportional to the probability that the animal chosen was \( D^+ \). Calculation of the approximation is still complex (involving a summation over \( n \) terms) but is much faster than calculation of the exact hypergeometric probability.

2.5.2. Performance of the approximation

While in most situations the approximation provides a result almost identical to the true probability, the difference can be important in some circumstances. The formula was evaluated by comparing it to the true probability (as obtained from the modified hypergeometric Eq. (8) and also to the probability assuming an infinite population (using the binomial Eq. (4))). Performance was evaluated over a range of population sizes (\( N \)) from 5 to 150, and over a range of sample sizes (\( n \) from 5 to \( N \)). For each combination of \( N \) and \( n \), the probability of detecting a number \( x \) of \( T^+ \) animals from a population with a prevalence of 0.2 using a test with sensitivity 0.95 and specificity 0.95 was calculated over all possible values of \( x \) (from zero to \( n \)). The difference between the true probability and the probability predicted by the formula was squared, and the mean of these squared differences taken over the \( n + 1 \) values of \( x \). The square root of this mean, the root mean square error (RMSE), was recorded for each combination of \( N \) and \( n \). RMSE is a measure of the average performance of the formula over the entire probability distribution, and can be expressed in the same terms as the probability (i.e. as a proportion or percentage).
The 11,026 data points produced were analysed by fitting an error surface to the three-dimensional data (\(N\) on the \(x\) axis, \(n\) (expressed as a percentage of \(N\)) on the \(y\) axis, and RMSE in the \(z\) axis) using a distance-weighted least-squares smoothing procedure (Anon., 1995). Fig. 1 shows the error-surface contours for the approximate formula and Fig. 2 shows the error surface contours for the simple binomial (infinite population) formula for comparison.

These graphs may be used to decide whether the use of the approximation is adequate for a particular purpose. The 0.01 contour represents the combinations of population size and sample size (as a percentage of the population size), that yield an average error in the probability estimate of 1%. If this level of accuracy is deemed to be the maximum error permissible, the approximation can be safely used with population and sample size combinations falling in the area below the 1% contour. For example, one may use sample sizes of less than 60% of the population (for small populations of less than 30), or sample sizes of less than 80% of the population for populations of 60 or above. For larger populations (greater than about 150), the formula yields accurate results when as many as 90% of the population are sampled (as may be necessary when examining herds with tests of relatively low sensitivity or specificity). By comparison, the binomial formula for infinite population sizes (Fig. 2) is, not surprisingly, decidedly inferior for population sizes in the range shown.

![Graph showing error-surface contours](image-url)

Fig. 1. Error-surface contour graph for the root mean square error of the probability distribution produced by the modified binomial approximation to the hypergeometric for a range of population and sample sizes. Root mean square error contour interval width = 0.005.
The results shown are valid only for a prevalence of 0.2, sensitivity of 0.95 and specificity of 0.95. The relationships between prevalence, sensitivity, specificity and RMSE are complex and difficult to generalise. Increasing prevalence may either increase or decrease RMSE depending on sensitivity and specificity. Varying sensitivity and specificity independently result in a maximum RMSE at either a sensitivity or specificity of 0.75. The error decreases more rapidly towards 0.5 than it does towards 1, and when a very high proportion of the population is sampled, the maximum RMSE occurs at a sensitivity or specificity of 1.

3. Sample size calculation

3.1. Establishing the hypothesis

When conducting a survey to prove freedom from disease, the null hypothesis, $H_0$ is that disease is present at a level equal to or greater than a specified prevalence. This prevalence may be chosen in two ways. For within-herd surveys, the prevalence
represents the minimum prevalence expected for a disease should it be present (highly contagious diseases, with long-lasting antibodies, might be expected to have very high minimum seroprevalences, while rare genetic diseases may have very low prevalences). Alternatively, the prevalence may be chosen to represent the population's level of disease that is small enough to be considered negligible or where, for infectious diseases, epidemic theory suggests it would move towards extinction even without intervention. For national-level surveys, the prevalence of disease-positive herds may be very low (regardless of the within-herd prevalence of positive herds). Choosing a minimum expected prevalence of, say, 1% of herds affected results in a survey that is unlikely to be able to detect a prevalence of < 1%. A survey indicating freedom from disease in such a case is, in fact, saying that if disease is present, it is present at a prevalence of < 1%.

The alternative hypothesis, $H_A$, is that the disease is present at a level lower than the specified prevalence. Thus:

$$H_0: \text{Prevalence} \geq p$$

$$H_A: \text{Prevalence} < p.$$
Fig. 3. Distributions showing the probability of observing a given number of reactors for a population with zero prevalence (dotted lines) and prevalence = 0.2 (solid lines), given two different sample sizes. Population size 100,000; sensitivity = specificity = 0.95.

At low sample sizes, the distributions (for the null and alternative hypotheses) have a wide overlap as can be seen from the left graph in Fig. 3 (graph a). As the sample size increases, the distributions become more separated, with a smaller overlap as shown on the right in Fig. 3 (graph b). The required sample size is the value at which the number of reactors at the cutpoint with a probability of $1 - \beta$ from the distribution with zero prevalence is equal to the number of reactors occurring with probability $\alpha$ at the left tail of the distribution with prevalence = $p$.

This procedure (described in detail in Appendix B) has been implemented in a computer program called FreeCalc. It was written in Borland Pascal 3 to run under MS DOS and includes sample-size and data-analysis modules. Features include a user-friendly mouse or keyboard interface and a comprehensive context-sensitive help system. All parameters used in calculations are able to be customised by the user, including $\beta$ and $\alpha$ (according to the significance of these two types of error) and the formula to be used in probability calculations (the exact hypergeometric (Eq. (8)), modified binomial approximation (described in Appendix A), or infinite population binomial formulae (Eq. (4))). The program is available free of charge over the Internet on the World Wide Web at the EpiVetNet Web site (http://epiweb.massey.ac.nz). No restriction is placed on the distribution of the program, and users are encouraged to pass it on to colleagues. Use of the program should always be acknowledged in reports, scientific papers and presentations.

---

4. Example

4.1. Foot and mouth disease

Let us assume that an outbreak of FMD has occurred in a previously free area. After the outbreak has been controlled, an effort is made to quickly show that the disease is no longer present. An unaffected herd of 265 animals within the control zone around the outbreak is to be examined two months after the outbreak to determine the presence of the disease. Let us assume that the test used is an ELISA with a sensitivity of 0.95 and specificity of 0.98. The highly infectious nature of the disease suggests that if the disease had passed through the herd, at least 30% of animals should have detectable titres (a more conservative approach would be to choose a lower expected herd prevalence).

The program can be used to calculate the sample size required to establish whether the herd is free from disease or not. The sensitivity (95%) and specificity (98%) of the test are entered, as well as the minimum expected prevalence (30%) and the population size (265). The values of \( \beta \) (the probability of falsely accepting the null hypothesis) and \( \alpha \) (the probability of falsely rejecting it) can be adjusted. In this case, let us assume that values of 0.05 are chosen for both. The exact hypergeometric probability formula is selected, and the ‘Calculate’ button pressed. After 7 iterations taking about 2 s (on a 486 33-MHz computer), a sample size of 14 is reported. The value of 14 was determined on the 6th iteration, with a reactor cutpoint level of 1, and the probability of observing 1 or fewer animals under the null hypothesis of 0.046. This tells the operator the following information: (i) if the herd is uninfected, the probability of observing \( \leq 1 \) reactors from a sample of 14 is equal to or just greater than 0.95 (power); and (ii) if the herd is infected with a prevalence of 30%, the probability of observing 1 or fewer reactors is equal to 0.046 (the closest value achieved \( \leq \alpha \)).

The modified binomial approximation formula produces a sample size of 14 as well, in about 1 s. The probability of observing \( \leq 1 \) animals under the null hypothesis is underestimated by \( 1.14 \times 10^{-5} \). By contrast, the assumption of a perfect test reduces the sample size to 9, an infinite population with the imperfect test produces the sample size of 14, and a perfect test and infinite population requires a sample of 9. A more conservative value of 10% for the minimum expected prevalence (instead of 30%) yields a sample size of 62 animals with a cutpoint number of reactors of 3. Assuming a perfect test reduces this to 27, an infinite population with the imperfect test yields 67, and a perfect test with infinite population requires a sample size of 29.

Let us assume that a sample of 62 animals is eventually taken from the herd (based on the conservative assumption of a minimum expected prevalence of 10%), and that 2 animals show positive reactions to the ELISA. The analysis module of the program can be used to help interpret this result. The sensitivity, specificity, minimum expected prevalence, and herd size are entered as above, followed by the sample size (62) and number of reactors (2). Pressing ‘Calculate’ produces the result that the probability of observing 2 or fewer reactors under the null hypothesis is equal to 0.013. We can therefore conclude with a high level of confidence that the herd is not diseased, and that the two reactors were probably due to the imperfect specificity of the test.
Finally, let us imagine that 5 animals tested positive instead of 2. Using the same approach, the analysis module tells us that the probability of observing 5 of fewer animals under the null hypothesis is 0.244. In this instance, the herd would be considered as infected until more information can be obtained.

Note that no attempt has been made to estimate the true prevalence based on the results of the survey. This is because the survey was not designed to make an estimate of the prevalence, but to determine how likely it was that disease was present in the herd.

4.2. *Paratuberculosis*

In contrast to FMD (a highly infectious disease for which good tests are available), paratuberculosis (Johne’s Disease) is a disease of low infectivity. Several tests are available, but one practical screening test is the absorbed ELISA. For example, let us assume that the sensitivity is 60% and specificity is 99% (Ridge et al., 1991). The minimum expected prevalence of this slow-spreading disease might be estimated at 2%. When calculating the sample size required to show freedom from paratuberculosis for a herd of the same size as before (265 animals), FreeCalc reports that it is unable to achieve the required accuracy even by sampling every unit. With a sample size of 265 (the whole herd), the probability of observing 6 reactors from a disease-free herd is 0.052, while the probability of observing 6 reactors from an infected herd with a prevalence of 2% is 0.695. The test is therefore unable to distinguish between an infected and uninfected herd at the 95% level of confidence as specified.

In larger herds, one approach to overcoming this problem is to increase the sample size. From a herd of 1000 animals, a sample of 989 is able to detect disease at the required confidence level. For smaller herds, this is not possible. The second approach is to lower the required power and/or level of confidence of the survey. Reducing the power from 95% to 72%, and the confidence level from 95% to 85% means that a sample of 259 animals from 265 will meet the requirements (again, this is not a very satisfactory solution). A third option is to relax (i.e. raise) the minimum expected or tolerable prevalence that the survey is required to detect. A sample of 237, with a cutpoint number of reactors of 5 is adequate to detect prevalences ≥5% in a herd of 265. While this is perhaps the practical limit of the accuracy of the ELISA, many herds can have a lower non-zero prevalence and be missed. The last option is to increase the sensitivity (modify the test, use a different test, or use tests in parallel). In this example, changing the critical value of the ELISA at which an animal is said to be infected may achieve the desired result. Let us assume that shifting the critical value produces a test with sensitivity of 72.9% and specificity of 84.8% (Spangler et al., 1992). The new form of the test is able to achieve the required level of precision with a sample size of 248 from a herd of 265, and a cutpoint number of reactors of 44.

This example serves to demonstrate that the use of tests of low sensitivity to attempt to substantiate freedom from diseases of low prevalence is extremely difficult. Even when practically achievable sample sizes are less than those recommended by the formula, the survey results may, by chance, be extreme enough to provide a definitive answer at the level of confidence required.
5. Discussion

The formula presented in this paper yields the exact probability of observing a given number of diseased animals, when sampling from a finite population with an imperfect test. We have also presented an approximation to this formula, which is computationally somewhat simpler, and shown the limits of accuracy for this approximation. These formulae have important application to the planning and analysis of surveys to substantiate freedom from disease. First, they may be used to calculate the precise sample size required for such surveys. Secondly, they enable the results of these surveys to be accurately analysed.

Unlike the formulae in common use, the new formulae are mathematically complex. While they provide a sound basis for the conduct of surveys for herd accreditation or substantiation of national disease status for international trade, their complexity could hamper the adoption of this approach by disease control authorities. The FreeCalc program was developed to overcome this problem.

An interesting and potentially important characteristic of probability formulae that consider finite populations (including the new formula described here) arises when the number of diseased animals in a population is expressed in terms of an estimate of prevalence. In small populations, prevalence can take only a finite number ($N$) of discrete values. When the estimated prevalence is used, $d$ must be calculated by rounding $pN$ to an integer value. Rounding down (truncating) gives the most conservative estimate. With increasing population size, $d$ increases in integer steps (with an interval equal to the inverse of $p$) rather than smoothly in proportion to $N$. The true prevalence $(d/N)$ varies between $p$ and slightly below $p$. This gives rise to the paradoxical situation where, although increasing $N$ may be expected to give an increasing probability of detecting a given number of fewer test positive animals from the population, sometimes increasing $N$ by 1 results in a fall in the probability. An awareness of the difference between estimated prevalence and the actual number of infected animals in a population of a given size will avoid confusion.

The use of the FreeCalc computer program enables a valid probabilistic approach to herd-level assessment. This will be of significant value to disease control authorities when large numbers of herds are to be tested, as is common in either accreditation schemes or national surveys to prove freedom from disease. This probabilistic approach means that costly and time consuming follow-up of positive reactors may not be necessary, and that the results of testing can be reported with more soundly based measures. This issue requires further research.

The program also accommodates the calculation of sample sizes for two-stage sampling. Most large-scale livestock surveys use a two-stage sampling scheme both to overcome problems with true random sampling from a large population, and to account for the clustering of disease within herds. When planning a two-stage sampling survey, sample sizes need to be calculated for both the first stage (selection of herds) and the second (selection of animals with selected herds). The use of the FreeCalc program to calculate cost-effective sample sizes for two-stage sampling surveys has been described (Cameron and Baldock, 1998).
Acknowledgements

This research was conducted as part of a project funded by the Australian Centre for International Agricultural Research, working in cooperation with the Thai Department of Livestock Development. Angus Cameron is supported by a research fellowship from the Australian Meat Research Corporation.

Appendix A. Derivation of binomial approximation

Fig. 4 shows the probability tree for selection from a finite population. If \( p_i \) is the probability of selection a disease-positive animal at the \( i \)th selection, when using a perfect diagnostic test in a finite population, \( p_1 = d/N \), \( p_2 = (d-1)/(N-1) \), \( (1 - p_{i+1}) = (N-d)/(N-x) \) and \( (1 - p_i) = (N-d-n+x+1)/(N-n+1) \). The probability of choosing \( x \) positive animals in any order, is given by the hypergeometric distribution. At each selection of a positive animal, the remaining number of positives is decreased by one, and the total number of animals is decreased by one. When a negative animal is selected, the remaining pool of negatives is decreased by one. The probability of selecting the next animal is then dependent on the remaining pool of positive and negative animals.

When using an imperfect test and the outcome of interest is not \( D^+ \) but \( T' \), the selected \( T' \) animal may be \( D^+ \) or \( D^- \). For a \( T' \), the probability that the animal is truly \( D' \) (positive predictive value) is given by the Bayes Rule:

\[
P(D'|T') = \frac{P(D')S_e}{P(D')S_e + P(D')(1 - S_p)}
\]

An approximation to the exact probability may be made by assuming that when a \( T' \) animal is selected, the pool of \( D^+ \) animals will be decreased by a ‘fraction of an animal’ equal to \( P(D^+|T') \). At the same time, the pool of \( D^- \) animals will be decreased by \( P(D^-|T') \), the complement of \( P(D^+|T') \). This will result in an overall decrease in \( N \) of one animal. In the same way, when a negative animal is selected, the pool of \( D^+ \) animals is decreased by \( P(D^+|T^-) \) and the \( D^- \) pool is decreased by \( P(D^-|T^-) \). This non-integer population of animals represents the mean expected population over many trials.

Unfortunately, the probability of drawing \( T^+ \) and then a \( T^- \) is not the same as the probability of drawing \( T^- \) then \( T^+ \) when this approach is used. The usual method of calculating the probability (first drawing all the \( T^+ \) followed by drawing \( n-x \) \( T^- \), and multiplying by the number of different possible combinations) will result in a biased estimate. An approximation that can be used to overcome this problem is to calculate the probability of drawing \( x \) \( T^+ \) based on the ‘average’ combination. This represents the combination where the \( T^+ \) ‘s are spread evenly through the \( n \) animals. Multiplying the probability of the average combination by the number of possible combinations yields a good estimate of the sum of all combinations.
The formula is most practically calculated using a computer program that keeps a tally of the remaining pool of $T^+$ and $T^-$. If `combination` is an array of $n$ boolean values (with $x$ true values, and $n-x$ false values regularly spaced) representing the 'average' combination of $x$ $T^+$ and $n-x$ $T^-$, $pos$ is the initial number of diseased animals in the population, $neg$ is the initial number of non-diseased animals, and $sens$ and $spec$ are sensitivity and specificity of the test, the following pseudo-code may be used for the calculation.

```
for counter = 1 to $n$
  if combination[counter] is true then
    Prob1 = (pos/(pos + neg)) × sens  {the animal tests positive}
    Prob2 = (neg/(pos + neg)) × (1 - spec)
  else
    Prob1 = (pos/(pos + neg)) × (1 - sens)
    Prob2 = (neg/(pos + neg)) × spec
  endif
  probability = probability × (Prob1 + Prob2)  {cumulative product}
  pos = pos - (Prob1 + Prob2)  {decrease the pool of $T^+$}
  neg = neg - (Prob2/(Prob1 + Prob2))  {decrease the pool of $T^-$}
end
probability = $n \choose x$ × probability  {multiply by number of combinations}
```

**Appendix B. Sample size calculation**

A trial and error procedure is used to calculate the required sample size. An arbitrary starting sample size (50) is chosen, and the probability of falsely concluding that disease is not present is calculated. If this probability is higher than the required level of $\alpha$, the sample size is increased, and the calculation starts again. If it is lower than the required level of $\alpha$, the sample size is decreased. The process is continued until a value equal to or just smaller than $\alpha$ is reached.

The cutpoint for a certain sample size represents the maximum number of reactors that could be observed with $1 - \beta$ probability if the population is disease-free. This is calculated by first determining the probability of observing no reactors. If this probability is less than $1 - \beta$, the cumulative probability of observing 1 or fewer reactors is calculated. This is repeated until the probability is equal to or slightly greater than $1 - \beta$.

The cumulative probability of observing a number of reactors equal to or less than the cutpoint is calculated by summing the probabilities of observing 0, 1, 2... reactors up to the cutpoint, based on a population with disease prevalence as set by the null hypothesis. This cumulative probability is compared with $\alpha$ to assess whether the sample size is correct, too large or too small.

The sample size is calculated on the basis of a cutpoint which is chosen to achieve a specified value of $\beta$. This cutpoint is calculated separately for each sample size,
considering the population size, test sensitivity and specificity. Because cutpoints can only increase by integer values, the probability curve follows a saw tooth shape as shown in Fig. 5. Although an increase in sample size usually leads to a decrease in probability, when there is a step in the cutpoint, an increase in sample size leads to a sharp increase in probability. In the example shown in Fig. 5, if a value of 0.1 had been chosen for $\alpha$, there are 3 points (marked with arrows) at which the probability curve falls below the $\alpha$ value. The point with the lowest sample size is the one required, but the trial and error search technique may arrive at any of these points.

To overcome this problem and identify the first point with the lowest sample size, the cutpoint is reduced by one, and the new sample size calculated (corresponding to the low-point of the previous saw-tooth in Fig. 5). If the probability for this lower sample size is less than the $\alpha$ value, the process is repeated, until the lowest cutpoint yielding a probability equal to or less than $\alpha$ is found. Finally, the sample size is progressively reduced by one, until the lowest sample size is found (moving up the selected saw-tooth until it intersects with the $\alpha$ probability level of 0.1).

References


Comparison of the Johne’s absorbed EIA and the complement-fixation test for the diagnosis of Johne’s
Spangler, C., Bech-Nielsen, S., Heider, L.E., Dorn, C.R., 1992. Interpretation of an enzyme-linked immuno-
sorbent test using different cut-offs between positive and negative samples for diagnosis of paratuberculo-
Surveillance 22, 22–24.