Estimation of animal-level prevalence from testing of pooled samples

Report prepared for

The Australian Biosecurity Cooperative Research Centre for Emerging Infectious Diseases

by

AusVet Animal Health Services
(Evan Sergeant)

and

The University of Sydney
(Jenny-Ann Toribio)

31 December 2004

AusVet Animal Health Services Pty Ltd
PO Box 3180, South Brisbane Qld 4101
Tel: 07 3255 1712 Fax: 07 3844 5501
www.ausvet.com.au

Contents

Executive Summary

This project has supported the development and testing of an internet-based calculator for the estimation of animal-level prevalence based on the results of testing of pooled samples. Testing of the calculator and case studies have shown that these methods have application in a wide range of situations and produce results with accuracy and precision that are comparable to unpooled (individual) testing. Pooled testing also provides potential for considerable savings in the total number of tests required. Alternatively, pooled testing can provide much greater precision of estimates for the same number of tests, at the cost of an increase in the total sample size.

The Pooled Prevalence Calculator is particularly useful for both survey design and analysis using pooled testing. It can be accessed via the internet, at http://www.ausvet.com.au/pprev/. To assist with survey design, it includes sample size calculators to estimate the required sample size for various pooling levels for both perfect and imperfect test sensitivity and specificity. In addition, a simulation module allows further evaluation and comparison of sampling and pooling strategies by simulating sampling and testing. The preferred strategy can then be selected to minimise bias in the resulting prevalence estimate, while at the same time meeting other survey design requirements.

Once the survey has been completed, the results can be analysed to provide realistic and precise prevalence estimates. Several alternative methods are provided to suit a range of situations, depending on whether pool size was fixed or variable and whether test sensitivity and specificity were assumed to be 100%, fixed values less than 100% or whether their true values were uncertain. The calculator can also be used to support demonstration of freedom from disease if all pools produce a negative test result. It will calculate an upper limit for true prevalence, with a specified level of confidence, if all pools test negatively and can also calculate the required sample size to provide the specified level of confidence that true prevalence is less than the (specified) design prevalence.

Pooled testing does have the disadvantage that estimates can be biased, particularly if prevalence is high (>20-30%) or pool size is large. In these situations, pooling might have only a marginal benefit over unpooled testing, because of the requirement for small pool sizes and large numbers of pools. It is also essential that samples are selected and allocated to pools in a random manner to prevent clustering and resultant biasing of results.

Based on the results of case studies, pooled testing has widespread potential application in surveillance for new and emerging diseases. Large numbers of samples can be collected and pooled for subsequent testing and analysis of results. If all pools are negative, there is strong evidence for freedom from the disease, while if some pools are positive it is possible to estimate prevalence with a high level of precision. Simulation can be used to develop pooling strategies suited to particular circumstances, to ensure that optimum pool sizes are used, to minimise bias and provide a reliable result. Using pooled testing in this way has the potential for significant cost-savings without sacrificing the validity or precision of the resulting prevalence estimates, particularly for diseases that are expected to be at low to moderate prevalence.

1. Introduction

Pooled or group testing is a testing strategy where samples from a number of individuals are aggregated into a single sample (or pool) which is then tested for the disease or agent of interest. Pooled testing strategies have been proposed for identification of infected individuals (classification), for demonstration of herd or area infection status and for estimation of prevalence of infection (Dorfman R, 1943; Worlund DD and Taylor G, 1983; Chen CL and Swallow WH, 1990; Kacena et al., 1998; Jordan D, 2003).

When used for classification testing, all individuals from negative pools are assumed to be free of disease, while individuals from positive pools are retested, either individually or in smaller pools until either a negative (pooled or individual) or positive (individual) result is achieved for each individual. This approach requires that there is sufficient sample from each individual to allow the required amount of retesting to be undertaken, and is therefore not appropriate where each sample can only be tested once.

When used for herd or area classification, sample size is determined to provide the desired confidence of detecting infection if it is present at a specified (design) prevalence. Samples are then tested in pools and if all pools are negative the herd or area is assumed to be free of disease at the specified prevalence, while if one or more pools are positive the disease is demonstrated to be present.

When used for estimation, individual samples are aggregated into pools for testing. Individual-level prevalence is then estimated based on the number of individuals represented in each pool and the test result for that pool. Estimation methods have been developed for both fixed and variable pool size and to adjust estimates for imperfect sensitivity and specificity of the test used.

Pooled testing approaches have been used in a wide variety of circumstances, including estimating infection rates in insect vectors, or for estimation of disease prevalence in animal, fish or human populations using a variety of test types (Thompson KH, 1962; Kerr JJ, 1971; Worlund DD and Taylor G, 1983; Kline RL et al., 1989; Kacena et al., 1998; Cowling et al., 1999a; Williams CJ and Moffitt CM, 2001). Probably the single disease that has led to the greatest interest in pooled testing strategies in the last two decades has been HIV-Aids, and the need for effective and economic strategies for classification of individuals and for estimation of prevalence (Kline RL et al., 1989; Tu et al., 1994; Mendoza-Blanco et al., 1996; Verstraeten et al., 1998; Brookmeyer, 1999).

Pooled testing has some significant advantages over individual testing. In particular, pooled testing provides:

- significant cost savings, particularly where the cost of collecting additional samples is small relative to the cost of the test or if prevalence is low (Sobel M and Elashoff RM, 1975; Chen CL and Swallow WH, 1990; Muñoz-Zanzi CA et al., 2000);
- increased precision of estimates compared to individual testing where the same number of tests (pools) is undertaken, particularly if prevalence is less than about 30% (Kline RL et al., 1989; Worlund DD and Taylor G, 1983);
- increased precision of estimates compared to individual testing where test sensitivity and/or specificity are less than 1 (Abel et al., 1999); and

- reduced bias in estimates when assumed values for sensitivity and specificity are not equal to the true values (Abel et al., 1999).

Pooled testing strategies also have a number of disadvantages, including:

- logistical requirements and cost for pooling of samples, either in the laboratory or in the field;
- the requirement of some strategies for re-testing of samples from positive pools either as smaller pools or individually, increasing requirements for sample volumes, storage and handling (Hughes-Oliver JM and Swallow WH, 1994; Chen CL and Swallow WH, 1990);
- the possible effect of dilution on test performance, particularly test sensitivity cannot be ignored, and any adjustment for imperfect sensitivity should include consideration of any effect of dilution on test sensitivity (Worlund DD and Taylor G, 1983); and
- estimates may be biased to a variable degree, depending on prevalence, number of pools and pool size (Thompson KH, 1962; Worlund DD and Taylor G, 1983; Tu et al., 1994; Abel et al., 1999).

The methods depend on a number of major assumptions, and may result in biased estimates if the assumptions are violated. These assumptions include that:

- the outcome is assumed to follow a binomial distribution – clustering or overdispersion of the positive outcome can result in substantial bias in the resulting estimate (Thompson KH, 1962; Chen CL and Swallow WH, 1990);
- sample size is small compared to the population (Thompson KH, 1962);
- the health status of each individual is independent of the status of others, both within and between pools (Worlund DD and Taylor G, 1983);
- the assumed values for test sensitivity and specificity are correct (Cowling et al., 1999a; Williams CJ and Moffitt CM, 2001);
- samples are assumed to be mixed homogeneously in the pools and any sub-samples taken for testing are equally representative of all of the individuals contributing to each pool (Jordan?); and
- all pools represent the same number of individuals (except for the variable pool-size method).

This review will concentrate on pooled testing strategies and analytical methods for the estimation of disease (or other characteristic) prevalence in individuals. In particular the review will focus on methods for estimation of prevalence suitable for implementation in an epidemiological calculator to support their wider use. Methods reviewed include frequentist methods for fixed or variable pool sizes and for assumed perfect tests, tests of imperfect but known sensitivity and specificity and for imperfect and uncertain sensitivity and specificity of tests. Bayesian methods for estimating true prevalence of disease from pooled sampling with fixed pool sizes are also discussed as well as related Bayesian methods for estimating true prevalence from individual (unpooled) testing using one or two tests with imperfect sensitivity and/or specificity.

2. Estimating prevalence from pooled samples

Except where indicated otherwise, the notation used for all methods described is as shown below.

Variable	Description
m	Number of pools tested
k	Pool size (number of individuals represented in each pool)
n	Total number of individuals sampled = k × m
x	The number of pools that have a positive test result
p	Estimated prevalence of the characteristic among individuals
e	Desired precision (= acceptable error) of the prevalence estimate
π	True prevalence of the characteristic among individuals
P	The proportion of positive pools = x/m
Se	Test sensitivity
Sp	Test specificity
Var(p)	Variance of p
$Z_{\alpha/2}$	The standard normal variate for calculating the 100(1 - α)% confidence limits

2.1. Frequentist methods for fixed pool sizes

A number of methods have been developed for estimating prevalence where all pools are the same size, and were recently reviewed by Cowling et al (1999). These methods assume that the number of test positive groups is binomially distributed and use maximum likelihood methods to estimate the true prevalence at the individual level. Methods have been developed for:

1. Assumed perfect tests with asymptotic confidence limits;
2. Assumed perfect tests with exact confidence limits;
3. Tests of known sensitivity and specificity with asymptotic confidence limits;
4. Tests of known sensitivity and specificity with exact confidence limits; and
5. Tests of uncertain sensitivity and specificity with asymptotic confidence limits.

Method 1: Perfect tests with asymptotic confidence limits

If the prevalence of a characteristic in a population is π and a series of equally sized pools (of size = k) are tested, the probability that any pool will have a negative test result is $(1 - \pi)^k$. From this, the maximum likelihood estimate of π can be calculated from Equation 1 and the variance of p can be calculated from Equation 2 (Thompson KH, 1962; Kerr JJ, 1971; Griffiths, 1972; Kline RL et al., 1989; Cowling et al., 1999a).

$$p = 1 - (1 - x/m)^{1/k} \quad (1)$$

$$Var(p) = ((x/m) \times (1 - x/m)^{2/k - 1})/(mk^2) \quad (2)$$

An alternative method for calculating the variance of p is (Sacks JM et al., 1989; Worlund DD and Taylor G, 1983):

$$Var(p) = (1 - p)^2 \times ((1 - p)^{-k} - 1)/(mk^2) \quad (3)$$

Asymptotic (approximate) confidence limits for the prevalence can be calculated as (Kline RL et al., 1989; Cowling et al., 1999a):

$$CL_{\alpha/2} = p \pm Z_{\alpha/2} \times \sqrt{Var(p)} \quad (4)$$

Because these asymptotic confidence limits are based on large-sample theory, the lower confidence limit can be negative when prevalence is close to zero (Cowling et al., 1999a). This method also assumes that the test used is 100% sensitive and specific, so that p will be biased if imperfect tests are used.

Method 2: Perfect tests with exact confidence limits

For this method, p is estimated in exactly the same manner as for Method 1, using Equation 1. To overcome the problem of negative lower confidence limits, a method was developed to calculate exact confidence limits. Briefly, exact confidence limits are obtained for the proportion of positive pools and then transformed to individual-level prevalence values using Equation 1 (Hauck WW, 1991; Cowling et al., 1999a). This method also allows calculation of an upper confidence limit where all pools test negative (Hauck WW, 1991). See Cowling et al. (1999) (Method 3) for details of the calculations.

Method 3: Tests of known sensitivity and specificity with asymptotic confidence limits

Prevalence estimates that assume perfect test sensitivity and specificity can be seriously biased if these assumptions are not valid (Rogan and Gladen, 1978). The direction and magnitude of the bias depend on the true values of sensitivity, specificity and prevalence. The above methods both assume the use of a perfect test and therefore resulting prevalence estimates will be biased if imperfect tests are used. To overcome this limitation, Tu et al. (1994) proposed an adjustment for imperfect sensitivity and specificity of tests, as shown in Equation 5 (Tu et al., 1994; Cowling et al., 1999a).

In this context, sensitivity and specificity are estimated at the pool level. For example, sensitivity is the probability that a pool containing samples from one or more infected individuals will produce a positive result in the pooled test. Pool-level sensitivity is therefore affected by both the prevalence of disease and pool size. The higher the prevalence, the more infected individuals that will be represented in individual pools and the more likely a pool is to test positive and therefore the higher the sensitivity. This is in contrast to individual-level sensitivity, which is independent of prevalence. However, the larger the pool size, the greater the dilution of any positive individual samples, potentially reducing sensitivity. Conventional individual-level sensitivity and specificity values can be used if there is evidence that there is no change in these values due to dilution for proposed pooling rates (Tu et al., 1994).

$$p = 1 - ((Se - x/m)/(Se + Sp - 1))^{1/k} \quad (5)$$

This method is constrained by the values of Se and Sp, such that $(1 - Sp) \leq x/m \leq Se$(Tu et al., 1994). For values of x/m outside this range p cannot be calculated.

The variance of p can be calculated using Equation 6 and asymptotic confidence limits can then be calculated using Equation 4 (Tu et al., 1994; Cowling et al., 1999a).

$$Var(p) = ((1 - p)^{2/k - 2}/mk^2) \times ((x/m \times (1 - x/m))/(Se + Sp - 1)^2) \quad (6)$$

Method 4: Tests of known sensitivity and specificity with exact confidence limits

As for Method 1, Method 3 can result in the lower confidence limit being negative if prevalence is low. Therefore, the approach used in Method 2 for calculation of exact confidence limits was extended to allow calculation of exact limits for imperfect tests of known sensitivity and specificity (Tu et al., 1994; Cowling et al., 1999a). This approach also

uses Equation 5 to calculate p, and substitution to calculate exact confidence limits. See Cowling et al. (1999) (Method 5) for details of the calculations.

Method 5: Tests of uncertain sensitivity and specificity with asymptotic confidence limits
Methods 3 and 4 assume that exact values for the sensitivity and specificity of the test are known (but < 100%). In most cases this is not the case, and sensitivity and specificity must be estimated from data (Cowling et al., 1999b). Thus, values of sensitivity and specificity are themselves uncertain, and introduce an additional source of error into the estimation of prevalence. In this situation, conventional confidence limits such as those calculated in Methods 3 and 4 underestimate the true uncertainty about the prevalence estimate and methods have been developed to support the use of tests with uncertain sensitivity and specificity (Rogan and Gladen, 1978; Cowling et al., 1999b).

For this approach, the point estimate of p is again the same as for Method 3 (Equation 5), using point estimates for test sensitivity and specificity. However, the variance of p (and therefore the resulting asymptotic confidence limits) is adjusted for additional uncertainty associated with the estimates of sensitivity and specificity as shown in Equation 7, where n_1 and n_2 are sample sizes used to calculate point estimates of sensitivity and specificity, respectively (Cowling et al., 1999a, Method 6).

$$\begin{aligned}\mathrm{Var}(p) = {} & (((Se - x/m)/(Se + Sp - 1))^{2/k-2})/(k^2 \times (Se + Sp - 1)^2) \\ & \times (x/m \times (1 - x/m)/m + (Se \times (1 - Se)/n_1) \\ & \times (1 - (Se - x/m)/(Se + Sp - 1))^2 \\ & + (Sp \times (1 - Sp)/n_2) \times ((Se - x/m)/(Se + Sp - 1))^2) \end{aligned} \quad (7)$$

A word about confidence limits
Asymptotic confidence limits for a binomial proportion, based on large-sample theory and the normal approximation, have been shown to provide quite poor and erratic coverage, and can often provide a quite misleading confidence in the result (Brown LD et al., 2001). Alternative methods have been proposed for confidence limits for simple binomial proportions, although these methods are not all applicable to pooled estimation methods. Because of this limitation, exact binomial confidence limits (where possible) should be preferred to the asymptotic alternative.

2.2. Frequentist approach for variable pool sizes

The above approaches all assume that pool size is the same for all pools tested. However, it may not always be possible or feasible to ensure that all pools are of equal size. In addition, several studies have shown that the precision of estimates can be improved by careful application of varying pool sizes (Hepworth G, 1996; Williams CJ and Moffitt CM, 2001). Various approaches have been suggested to accommodate alternative testing strategies with multiple pools of varying size (Walter et al., 1980; Farrington, 1992; Hepworth G, 1996; Brookmeyer, 1999; Williams CJ and Moffitt CM, 2001).

Variable pool-size methods have particular application in situations where:

- it is not always feasible to have exactly the same number of individuals per pool (for example when testing groups of insect vectors or when sample size is not exactly divisible into equal size pools) (Walter et al., 1980);

- a sequential testing approach is used (where additional pools of different sizes are tested depending on the results of preliminary testing on a small number of pools) (Hughes-Oliver JM and Swallow WH, 1994; Hepworth G, 1996);
- multi-stage sampling is used, where positive pools are subdivided into smaller pools and retested (Brookmeyer, 1999); or
- increased precision of the estimate is required where prevalence is relatively high (Williams CJ and Moffitt CM, 2001)

Briefly, these approaches utilise generalised linear models to calculate the maximum likelihood estimate and confidence limits for p, assuming perfect test sensitivity and specificity. See Farrington (1992) and Williams and Moffitt (2001) for details of the methods used.

2.3. Bayesian approach for fixed pool sizes

Bayesian analysis provides an alternative method for estimating prevalence from the results of pooled testing. Briefly, this approach allows the combination of any prior information available on test sensitivity and specificity and estimated prevalence of disease with the results of testing, to produce a posterior probability distribution of the estimated true prevalence. Bayesian methods were initially developed for estimating prevalence from individual testing (see Section 3 for more details – Joseph et al., 1995b; Enøe et al., 2000) and were subsequently extended for use with pooled testing strategies (Mendoza-Blanco et al., 1996; Cowling et al., 1999a).

Bayesian methods also have the advantages that:

- the lower probability limit can never be negative;
- imperfect sensitivity and specificity of tests and uncertainty about their true values are incorporated explicitly in the procedure; and
- any pre-existing estimates of prevalence can be incorporated in the analysis to increase confidence in the results

However, Bayesian estimates can be seriously affected by the use of inappropriate prior distributions (inaccurate estimates and/or overconfidence in the values) for prevalence, sensitivity or specificity and therefore must be used with care. Wherever possible prior estimates should be based on real data and should be appropriately weighted (wide probability limits) to ensure that any errors do not dominate the data, causing inaccurate results. Unless there is very good data (or expert knowledge) on which to base a prior distribution for prevalence (and other parameters) it is wise to use an uninformative (uniform) prior distribution. This is a distribution where all values between 0 and 1 can occur with equal probability, and indicates no prior knowledge about the true value of the parameter.

This method requires input of prior estimates for test sensitivity and specificity and disease prevalence as Beta probability distributions. The prior estimates of sensitivity, specificity and prevalence are estimates of the most likely values of these parameters, based on previous studies, alternative sources of data or expert opinion. Beta distributions are used because these distributions are commonly used to describe uncertainty about the true value of a proportion, such as sensitivity or prevalence, and are appropriate distributions for use in the Gibbs sampler (Joseph et al., 1995b; Vose, 2000). When used for this purpose, the Beta distribution can be defined by the two parameters, α and β, with $\alpha = x + 1$ and $\beta = n - x + 1$, where x is

the number of positive events out of n trials. As n increases, the degree of uncertainty (the width of the distribution) about the estimated proportion (x/n) decreases. Alternatively, if the mode (most likely value) and 5^{th} or 95^{th} percentile for the desired prior distribution can be determined from expert opinion or other sources the appropriate α and β parameters for this distribution can be calculated (Suess EA et al., 2002). A uniform distribution (discussed above) indicates no prior knowledge and has parameters $\alpha = \beta = 1$.

Outputs from the Gibbs sampler are revised estimates of prevalence, test sensitivity and test specificity as posterior probability distributions. See Mendoza-Blanco et al. (1996) and Cowling et al. (1999) (Method 7) for more details of the methodology and implementation of the Gibbs sampler.

3. Bayesian methods for unpooled samples

As mentioned above, Bayesian methods for prevalence estimation were originally developed for individual sampling strategies, and subsequently adapted for pooled testing. Traditionally, true prevalence has been estimated using the "Rogan-Gladen" estimator to adjust apparent prevalence for imperfect test sensitivity and specificity (Rogan and Gladen, 1978). As mentioned above, conventional confidence limits underestimate the true uncertainty about this estimate unless the true values of sensitivity and specificity are known. Bayesian methods provide an alternative approach to estimating prevalence that allows for imperfect test performance and uncertainty about the true values of sensitivity and specificity. In addition, these methods can be used to generate revised estimates of test sensitivity and specificity as well as prevalence (Joseph et al., 1995b; Enøe et al., 2000; Johnson et al., 2001).

3.1. Bayesian method using a single test of uncertain sensitivity and specificity

Where a single test is applied to a random sample of individuals from a population, a Gibbs-sampling approach can be used to estimate the true prevalence of disease, after adjusting for imperfect test sensitivity and specificity in a similar manner to that for pooled samples (Joseph et al., 1995b).

For this method, alpha and beta parameters for prior probability distributions for prevalence, sensitivity and specificity are required, and starting values for the assumed numbers of true-positive and false-negative individuals. Outputs from the Gibbs sampler are revised estimates of prevalence, test sensitivity and test specificity as posterior probability distributions. See Joseph et al. (1995) for more details of the methodology and implementation of the Gibbs sampler.

3.2. Bayesian method using two tests of uncertain sensitivity and specificity

A Gibbs-sampling approach can also be used to estimate the true prevalence of disease where randomly selected individuals from a population are tested with two tests concurrently (Joseph et al., 1995b). This method assumes that all individuals are tested using both tests, and that the two tests are independent, conditional on disease status. This means that the test sensitivity and specificity of one test remains constant regardless of the result of the second test. If the tests are not independent, prevalence will be underestimated (Sergeant et al., 2002).

For this method, alpha and beta parameters for prior probability distributions for prevalence and for the sensitivity and specificity of both tests are required, as well as starting values for

the assumed numbers of truly infected individuals in each cell of the 2 x 2 table describing the test results. Outputs from the Gibbs sampler are revised estimates of prevalence and test sensitivity and test specificity of both tests as posterior probability distributions. See Joseph et al. (1995) for more details of the methodology and implementation of the Gibbs sampler for two tests.

3.3. Bayesian method using two tests of uncertain sensitivity and specificity in two separate populations

Johnson et al. (2001) demonstrated that the posterior distributions for sensitivity and specificity of both tests in the two test, one population model described above were only marginally different from the prior distributions. They suggested that a better approach for estimating these parameters would be to use both tests concurrently on animals from two populations with different prevalences.

This method is an extension of the two-test, one population model but requires an additional prior distribution and produces an additional posterior distribution for prevalence in the second population. See Johnson et al. (2001) for more details.

4. Survey design issues

4.1. p is a biased estimator of prevalence

For all of the pooling methods described, p is a biased estimator of π (usually upwardly biased). This bias can be estimated as (approximately) $((k-1) \times Var(p))/(2 \times (1 - \pi))$ (Tu et al., 1994; Cowling et al., 1999a). This bias is decreased with lower prevalence (π), increased numbers of pools (m), smaller pool size (k) and increased total sample size (n) or as the probability that all pools will test positive increases (Worlund DD and Taylor G, 1983; Abel et al., 1999; Cowling et al., 1999a).

Worlund and Taylor (1983) suggested that the bias was negligible for $m \geq 30$. However, even for quite small values for m and k the bias is also quite small, particularly if p is low (Thompson KH, 1962).

The estimated prevalence (p) is also sensitive to (and may be biased by) errors in the assumptions of perfect test sensitivity or specificity. p is particularly sensitive to errors in sensitivity as p increases and if k is too large (Chen CL and Swallow WH, 1990). Clustering or overdispersion of positive individuals in the sampled population can also result in substantial bias in prevalence estimates (Chen CL and Swallow WH, 1990).

4.2. Sample size

Pooled testing of samples has the potential to substantially reduce the cost of prevalence surveys for disease, particularly if the true prevalence is low (Sobel M and Elashoff RM, 1975; Chen CL and Swallow WH, 1990; Muñoz-Zanzi CA et al., 2000). However, because of the effect of pooling strategies on prevalence estimates and confidence limits, conventional sample size calculations for prevalence surveys are not applicable for pooled testing strategies.

For the frequentist methods it is possible to calculate approximate numbers of pools required for given values for pool size, estimated prevalence and desired confidence and precision of the estimate, by transforming the formulae for calculation of confidence limits (Worlund DD and Taylor G, 1983).

From Equations 3 and 4, the required number of pools (m) to estimate p with a perfect test and fixed pool size (k) can be calculated using Equation 9, where e is the desired precision (allowable error) of the estimate (Worlund DD and Taylor G, 1983). Methods for tests of known and uncertain sensitivity and specificity are provided in Equations 10 and 11 respectively.

$$m_{(\text{perfect test})} = (Z(1-\pi)/(ek))^2((1-\pi)^{-k} - 1) \quad (9)$$

$$m_{(\text{known Se \& Sp})} = P'(1-P')Z^2(1-\pi)^{2/k-2}/(ek(Se+Sp-1))^2 \quad (10)$$

where $P' = (1-(1-\pi)^k Se + (1-\pi)^k(1-Sp)$

$$m_{(\text{uncertain Se \& Sp})} = fP'(1-P')/(e^2/Z^2 - f(Se(1-Se)(1-g)^2/n_1 + Sp(1-Sp)g^2/n_2)) \quad (11)$$

where $g = (1-\pi)^k$ and $f = g^{2/k-2}/(k\,(Se+Sp-1))^2$

For Equation 9 (fixed pool size and perfect test), the optimum value of k can be calculated that minimises the variance and consequently minimises the number of pools requiring testing to achieve the desired confidence and precision. This optimum value for k depends on the prevalence and has been previously demonstrated to be approximately 1.6/p (Thompson KH, 1962; Griffiths, 1972). This equates to the pool size which results in an expected number of 1.6 infected individuals per pool (Sacks JM et al., 1989). Although this optimum value for k minimises the number of pools required for given confidence and precision, it can result in biased estimates if the number of pools tested is small.

4.3. Managing bias

As noted above, p is a biased estimator of the true prevalence, although this bias is minimal under some conditions. The actual bias in any estimate depends on the true prevalence, pool size and the number of pools. The likely size of the bias for any particular pooling strategy can be estimated using simulation methods (Thompson KH, 1962; Tu et al., 1994; Abel et al., 1999; Williams CJ and Moffitt CM, 2001). Bias can be minimised by ensuring an adequate total sample size, by testing a larger number of pools of smaller size, rather than *vice versa* or by testing several individual samples in addition to the pooled samples (Williams CJ and Moffitt CM, 2001).

5. Implementing a Pooled Prevalence Calculator

5.1. Description

The Pooled Prevalence Calculator was developed using the R statistical language and environment, running on an internet web-server under Unix. Input values for the various calculations are passed to an R script from the HTML web-page, and results are returned to the web browser and displayed in HTML format. The site includes many of the methods as described above, with the specific methods and utilities provided described in more detail

below. A web-based user guide and glossary are also provided, as well as detailed instructions for the use of each method.

5.2. Prevalence estimation methods

The calculator includes a variety of methods for estimating pooled prevalence, including for:

- Fixed pool size and perfect sensitivity and specificity with asymptotic or exact confidence limits (Methods 1 and 2 above);
- Fixed pool size and imperfect but known values for sensitivity and specificity with asymptotic or exact confidence limits (Methods 3 and 4 above);
- Fixed pool size and uncertain estimates for sensitivity and specificity with asymptotic confidence limits only (Method 5 above);
- Fixed pool size and uncertain estimates for sensitivity and specificity using a Bayesian approach; and
- Variable pool size and perfect sensitivity and specificity.

For the frequentist methods for fixed pool sizes (Methods 1 – 5), outputs include estimated prevalence and confidence limits for the specified number of positive pools, as well as tables and graphs of prevalence estimates and confidence limits for all possible test results. For the variable pool size method, outputs include estimated prevalence and confidence limits for the specified test results.

For the Bayesian method, prior estimates of sensitivity, specificity and prevalence are required as Beta probability distributions, and outputs are presented as posterior probability distributions for the three parameters. Frequency histograms and density curves of these parameters are provided and detailed simulation results for each parameter are saved in a temporary file and can be downloaded if desired.

5.3. Sample size calculations

Sample size calculators are also provided for assumed perfect tests, tests of known sensitivity and specificity and tests of uncertain sensitivity and specificity. Sample sizes are calculated by transposition of the formulae used for to calculate confidence limits for Methods 1, 3 and 5 above (see also Worlund DD and Taylor G, 1983). These calculators provide estimates of the required number of pools to be tested for various pool sizes to achieve required precision and confidence in the estimated prevalence, for an assumed level prevalence. Outputs from these programs include a table of pool sizes, the corresponding numbers of pools required to be tested, the total sample size and a graph of numbers of pools required against pool size.

5.4. Simulation of pooled sampling

As discussed above, the pooled prevalence estimate is actually an upwardly biased estimator of the true prevalence. The level of bias can be evaluated for any particular sampling strategy using simulation methods.

Therefore, simulation programs are provided to simulate sampling and prevalence estimation for up to 6 different pooling strategies for assumed values of prevalence and test sensitivity and specificity and for a specified level of confidence. Options are available for both fixed and variable pool size. The program runs multiple iterations of sampling and estimation and calculates the mean prevalence, confidence interval width and estimated bias across all iterations. By simulating alternative pooling strategies this utility allows the various strategies

to be evaluated and compared to determine the optimum strategy that will give the desired level of precision in the prevalence estimate and also minimise the level of bias in the prevalence estimate.

Outputs for each simulation are summarised across all iterations for each strategy entered and presented in a summary table. The main outputs are:

- mean prevalence estimate;
- minimum and maximum prevalence estimates;
- mean bias in the estimated prevalence;
- mean confidence interval width;
- mean standard error of the estimated prevalence;
- mean squared error of the estimated prevalence (square of the mean standard error plus the square of the mean bias);
- relative bias as a proportion of the mean estimated (apparent) prevalence (AP);
- relative bias as a proportion of the specified design (true) prevalence (TP);
- squared mean bias as a proportion of the mean squared error;
- the proportion of iterations producing a valid estimate of prevalence (one in which the estimated confidence intervals contain the true (design) value; and
- detailed results for all iterations and histograms for each strategy are saved for viewing or downloading if desired.

5.5. Demonstration of freedom

Pooled testing can also be used as an alternative to individual testing in surveys to detect disease or to demonstrate freedom from disease. In this situation, a 'design' (or target) prevalence is specified which is regarded as the minimum prevalence at which the disease is likely to occur, or the prevalence below which the disease is not of concern. The program calculates the probability that one or more positive results would be detected in the sample if the population was infected at or above the design prevalence. The method used also allows for use of a test with imperfect sensitivity.

If all samples test negatively, we can calculate the probability that the true prevalence is equal to or greater than the design prevalence. This can then be interpreted as a statement of our level of confidence that the disease (if present) is at a prevalence less than the specified design prevalence. Alternatively, we can set our desired level of confidence (for example, 95%) and determine the prevalence that will provide the specified level of confidence of detecting the disease if it was present at a prevalence equal to or above that value.

The method used assumes fixed pool sizes and known test sensitivity only (sensitivity may be 100%). Test specificity is assumed to be 100%, either intrinsically, or through follow-up of any positives to confirm or refute infection. This approach to specificity is consistent with use of the test for detection of disease or demonstration of absence of disease.

5.6. Additional utilities

Additional utilities are also provided for:

- the estimation of prevalence using a Bayesian approach for unpooled (individual) testing using a single test of uncertain sensitivity and specificity (Joseph et al., 1995a);

- the estimation of prevalence using a Bayesian approach for unpooled (individual) testing using two independent tests of uncertain sensitivity and specificity(Joseph et al., 1995a);
- the estimation of alpha and beta parameters for Beta probability distributions for specified mode and 5th or 95th percentiles (Suess EA et al., 2002); and
- the calculation of summary values for Beta probability distributions for specified alpha and beta parameters.

6. Validation of results

Analyses were undertaken to demonstrate the methods and to validate them by comparison of results with previously published results (where available). Where previously published results were not available methods were validated by comparison with simulation results.

6.1. Frequentist methods for fixed pool size

The five frequentist methods implemented for fixed pool size (Methods 1 – 5) were validated by comparison of results with those of original estimates by Cowling et al. (1999) (provided by IA Gardner, UC Davis). Results were also compared with those of a spreadsheet implementing the various formulae for the range of possible outcomes for the given pool size and numbers of pools tested. Estimated prevalence, standard error and confidence limits were identical to four decimal places for all of the comparisons undertaken for the five methods.

6.2. Frequentist method for variable pool size

The method implemented for variable pool size was validated by analysis of data from Cowling et al. (1999), adapted to include varying pool sizes (700 pools of 20 and 20 pools of 19 instead of 719 pools of 20). Additional comparisons were also undertaken with results from previous analyses by Williams and Moffitt (2001) using examples with single pool sizes. Results from the variable pool size calculator were very similar to the results obtained using alternative methods.

6.3. Bayesian method for fixed pool size

The Bayesian method for estimating prevalence from pooled samples was validated by comparison of the posterior distribution for prevalence with results from Cowling et al. (1999) and with frequentist methods for uncertain test sensitivity and specificity. Results from the various methods were all very similar.

6.4. Bayesian prevalence estimation for unpooled samples

The Bayesian methods for estimating prevalence from unpooled testing using one or two tests were validated by comparison of results with original analyses by Joseph et al. (1996). Results were comparable to the published results.

6.5. Sample size estimation

Sample size estimation methods were validated by simulation to evaluate the precision and bias for the suggested pool sizes and numbers of pools for a range of scenarios. For all scenarios considered, the suggested combinations of pool size and number of pools resulted in prevalence estimates with the specified precision and minimal bias, except when options with large pool sizes and small numbers of pools were used.

6.6. Simulation of pooled sampling

The simulation methods could not be validated externally, because there was no available alternative data or method for validation. However, the results of all test simulations run were consistent with expectations, with estimates having minimal bias and expected precision for small pool sizes and large numbers of pools, and with increasing bias as pool size increased and number of pools decreased.

6.7. Demonstration of freedom using pooled sampling

Sample sizes (pool size and number of pools) and confidence levels for demonstration of freedom were calculated for pooled sampling were comparable to corresponding values calculated for unpooled sampling for a range of scenarios.

7. Demonstration analyses and case studies

In order to test out the calculator and to demonstrate its capabilities and utility, a number of demonstration analyses were undertaken as well as several case studies. These analyses are described in more detail in the Appendices and are summarised here:

7.1. Demonstration analyses

Demonstration analyses were undertaken for all the utilities provided in the Pooled Prevalence Calculator. For most analyses, a moderate-prevalence (14%) scenario for Hendra-virus testing in fruit bats was assumed. These demonstration analyses are described in Appendix 1 and are also available on-line by clicking the appropriate link on each data-input page.

7.2. Hendra & Nipah virus case studies

Case studies on sample size estimation, simulation of pooling strategies and prevalence estimation were also undertaken for three scenarios based on real (unpooled) surveillance data. The scenarios investigated were high (60%) and moderate-prevalence (14%) scenarios for Hendra virus in fruit bats and low (1.6%) prevalence for Nipah virus in Malaysian abattoir workers. The original data was based on individual (unpooled) testing of surveillance samples. The case studies evaluated the sample size and pooling strategies required to achieve similar results using pooled testing.

These case studies demonstrated that pooling of samples allowed a substantial reduction in the number of tests required without sacrificing accuracy or precision of the estimate. However, they also highlighted a slight increase in the overall sample size to achieve the same precision with pooled testing as for unpooled and the importance of selecting an appropriate pool size depending on the likely prevalence. For example, for the high-prevalence scenario, pool sizes greater than 2 resulted in substantial increases in overall sample size and increasing bias in the resulting estimate, whereas pool sizes of 5-10 and 20 or more were acceptable for the moderate and low-prevalence scenarios, respectively.

These case studies are described in more detail in Appendix 2.

7.3. Johne's disease case study

A separate case study was also undertaken using data or ovine Johne's testing. This data was from pooled faecal culture testing of sheep in a field trial of the effectiveness of vaccination

for the control of ovine Johne's disease. As well as pooled culture, samples from sheep contributing to positive pools (and some sheep from negative pools) were also cultured individually, allowing estimation of animal-level prevalence from unpooled testing. The case study compared pooled prevalence estimates, using several of the methods provided, with unpooled estimates for vaccinated and control (unvaccinated) cohorts at 30 months post-vaccination on three separate properties. This case study demonstrated that the pooled-prevalence estimation methods provided reasonable estimates of the estimated prevalence based on individual testing and, with only one exception, that the unpooled estimate was within the 95% confidence interval for the pooled estimate for all of the pooled methods and situations evaluated.

However none of these methods was ideal to address the complexity of the OJD trial data and the report concluded that further research was needed to develop a method that will account for variation in pool size and in test sensitivity and specificity at different pool sizes and prevalence levels. More investigations are also required to determine the best method for identifying an appropriate prior distribution for prevalence in situations such as this.

These case studies are described in more detail in Appendix 2.

8. Conclusion

Testing and case studies have shown that the Pooled Prevalence Calculator has application in a wide range of situations and produces results with accuracy and precision that are comparable to unpooled (individual) testing, at a considerable saving in the total number of tests required. Alternatively, pooled testing can provide much greater precision for the same number of tests, at the cost of an increase in the total sample size.

Although these methods have been recognised and discussed in the scientific literature for many years, this is the first time that they have been brought together and incorporated into a simple, user-friendly calculator and made readily available to researchers. Previously, methods had to be interpreted from the literature and implemented in spreadsheets or other computer software before they could be used by researchers. Some methods were sufficiently complex to discourage their use because of the difficulty in implementing in a simple and easy-to-use manner. This calculator therefore provides a useful addition to the tool-chest for researchers and epidemiologists involved in estimating prevalence for emerging and other diseases.

Extensive testing and evaluation has shown that the results produced by the calculator are accurate and can be used with confidence. In one instance (Johne's disease Case study), comparison with individual testing of samples showed that the 95% confidence intervals for pooled estimates consistently included the estimated prevalence from individual testing. In other examples, results were the same as previously reported results for the same method. Finally, simulation studies showed that the 95% confidence interval for the estimated prevalence included the true (design) prevalence in more than 95% of iterations, except when prevalence and pool size were high, resulting in biased estimates.

Case studies of Hendra and Nipah viruses and Johne's disease were undertaken to evaluate the reliability of pooled prevalence estimates in comparison to unpooled testing in real situations. These case studies included a wide range of scenarios, with estimated true prevalence based on individual testing, ranging from 0.6% to 59%. Based on the results of

these case studies, pooled testing has widespread potential application in surveillance for new and emerging diseases. In addition, simulation can be used to develop pooling strategies suited to particular circumstances, to ensure that optimum pool sizes are used to minimise bias and provide a reliable result. Using pooled testing in this way has the potential for significant cost-savings without sacrificing the validity or precision of the resulting prevalence estimates, particularly for diseases that are expected to be at low to moderate prevalence. Finally, if large numbers of samples are collected and pooled for subsequent testing and all pools are negative, there is strong evidence for freedom from the disease. In this situation the calculator can be used to calculate the level of confidence that prevalence is less than a specified design prevalence (or to calculate the upper confidence limit for true prevalence).

9. References

Abel U, Schosser R, Suss J, 1999. Estimating the prevalence of infectious agents using pooled samples: biometrical considerations. *Zentralblatt fur Bakteriologie* 289: 550-63.

Brookmeyer R, 1999. Analysis of multistage pooling studies of biological specimens for estimating disease incidence and prevalence. *Biometrics* 55: 608-12.

Brown LD, Cai T, DasGupta A, 2001. Interval estimation for a binomial proportion. *Statistical Science* 16: 101-133.

Chen CL, Swallow WH, 1990. Using group testing to estimate a proportion, and to test the binomial model. *Biometrics* 46: 1035-1046.

Cowling DW, Gardner IA, Johnson WO, 1999a. Comparison of methods for estimation of individual-level prevalence based on pooled samples. *Prev. Vet. Med.* 39: 211-25.

Dorfman R, 1943. The detection of defective members of large populations. *Annals of Mathematical Statistics* 14: 436-440.

Enøe C, Georgiadis MP, Johnson WO, 2000. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.* 45: 61-81.

Farrington CP, 1992. Estimating prevalence by group testing using generalized linear models. *Stat Med* 11: 1591-7.

Griffiths DA, 1972. A further note on the probability of disease transmission. *Biometrics* 28: 1133-9.

Hauck WW, 1991. Confidence intervals for seroprevalence determined from pooled sera. *Ann Epedemiol* 1: 277-281.

Hepworth G, 1996. Exact confidence intervals for proportions estimated by group testing. *Biometrics* 52: 1134-1146.

Hughes-Oliver JM, Swallow WH, 1994. A two-stage adaptive grouping procedure for estimating small proportions. *Journal of the American Statistical Association* 89: 982-993.

Johnson WO, Gastwirth JL, Pearson LM, 2001. Screening without a "gold standard": the Hui-Walter paradigm revisited. *Am. J. Epidemiol.* 153: 921-924.

Jordan D, 2003. A Simulation Model for Evaluating the sensitivity of pooled faecal specimens for classifying herd status of infection with fecal-borne pathogens. *Proceedings of the 10th International Symposium on Veterinary Epidemiology and Economics*. International Society for Veterinary Epidemiology and Economics,

Joseph L, Gyorkos TW, Coupal L, 1995a. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am. J. Epidemiol.* 141: 263-272.

Kacena KA, Quinn SB, Howell MR, Madico GE, Quinn TC, Gaydos CA, 1998. Pooling urine samples for ligase chain reaction screening for genital Chlamydia trachomatis infection in asymptomatic women.[see comment]. *J. Clin. Microbiol.* 36: 481-5.

Kerr JJ, 1971. The probability of disease transmission. *Bioetrics* 27: 219-222.

Kline RL, Brothers TA, Brookmeyer R, Zeger S, Quinn TC, 1989. Evaluation of human immunodeficiency virus seroprevalence surveys using pooled sera. *J. Clin. Microbiol.* 27: 1449-1452.

Mendoza-Blanco JR, Tu XM, Iyengar S, 1996. Bayesian inference on prevalence using a missing-data approach with simulation-based techniques: applications to HIV screening. *Stat Med* 15: 2161-76.

Muñoz-Zanzi CA, Johnson WO, Thurmond MC, Hietala SK, 2000. Pooled-sample testing as a herd-screening tool for detection of bovine viral diarrhea virus persistently infected cattle. *J Vet diagn Invest* 12: 195-203.

Rogan WJ, Gladen B, 1978. Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* 107: 71-76.

Sacks JM, Bolin S, Crowder SV, 1989. Prevalence estimation from pooled samples. *Am. J. Vet. Res.* 50: 205-206.

Sahani M, Parashar UD, Ali R, Das P, Lye MS et al., 2001. Nipah virus infection among abattoir workers in Malaysia 1998-1999. *Int. J. Epidemiol.* 30: 1017-1020.

Sergeant ESG, Whittington RJ, More SJ, 2002. Sensitivity and specificity of pooled faecal culture and serology as flock-screening tests for detection of ovine paratuberculosis in Australia. *Prev. Vet. Med.* 52: 199-211.

Sobel M, Elashoff RM, 1975. Group testing with a new goal, estimation. *Biometrika* 62: 181-193.

Suess EA, Gardner IA, Johnson WO, 2002. Hierarchical Bayesian model for prevalence inferences and determination of a country's status for an animal pathogen. *Prev. Vet. Med.* 55: 155-171.

Thompson KH, 1962. Estimation of the proportion of vectors in a natural population of insects. *Biometrics* 18: 569-578.

Tu XM, Litvak E, Pagano M, 1994. Studies of AIDS and HIV surveillance. Screening tests: can we get more by doing less? *Stat Med* 13: 1905-19.

Verstraeten T, Farah B, Duchateau L, Matu R, 1998. Pooling sera to reduce the cost of HIV surveillance: a feasibility study in a rural Kenyan district. *Tropical Medicine & International Health* 3: 747-50.

Vose D, 2000. *Risk Analysis – A quantitative guide*. 2nd edition. John Wiley and Sons Ltd., Chichester, England.

Walter SD, Hildreth SW, Beaty BJ, 1980. Estimation of infection rates in population of

organisms using pools of variable size . *Am. J. Epidemiol.* 112: 124-8.

Williams CJ, Moffitt CM, 2001. A critique of methods of sampling and reporting pathogens in populations of fish. *Journal of Aquatic Animal Health* 13: 300-309.

Worlund DD, Taylor G, 1983. Estimation of disease incidence in fish populations. *Canadian Journal of Fisheries and Aquatic Sciences* 40: 2194-2197.

10. Appendix 1: Demonstration analyses

10.1. Background

For these demonstration analyses (except where otherwise stated), a hypothetical pooled testing strategy for the estimation of prevalence of Hendra virus in fruit bats was used. The data was based on real data for Hendra virus testing in the little red flying fox (*Pteropus scapulatus*) in Queensland during the period 1996 – 1999 (H. Field, pers com). During this period, 162 samples were tested from little red flying foxes, with 22 samples positive, for an estimated prevalence of 13.6% (95% CI: 8.7 – 19.8%). For these analyses, pool sizes and numbers of pools were calculated to provide 95% confidence of estimating a true prevalence of 14 % with a desired precision of ± 5.5% for the various estimation methods, corresponding to the estimated prevalence and confidence interval for the original (unpooled) data. The most frequent result from simulation studies was then used to estimate the prevalence and confidence interval for each scenario.

10.2. Fixed pool size and perfect tests

Two methods are provided for fixed pool sizes and perfect tests, one producing asymptotic confidence intervals (Method 1) and the other exact confidence intervals (Method 2). These methods should only be used if you can be confident that the sensitivity and specificity of the test are both close to 100%. If the true prevalence is likely to be close to zero, Method 2 (with exact confidence limits) should be used in preference to Method 1, because Method 1 could produce a negative lower confidence limit.

For this analysis, it was assumed that samples from 210 individual fruit bats were aggregated into 42 pools of 5 samples each, that 22 pools produced a positive test result and that the test sensitivity and specificity were both 100%. Input values and results for this analysis are summarised in the table below.

	Method	
	1	2
Input values:		
Number of pools tested	42	42
Number of pools positive	22	22
Pool size	5	5
Lower CL	0.025	0.025
Upper CL	0.975	0.975
Results:		
Est. Prevalence	0.1379	0.1379
2.5 percentile	0.0866	0.0832
97.5 percentile	0.2038	0.1926
Std. Error	0.0279	0.0279

10.3. Fixed pool size and tests with known sensitivity and specificity

Two methods are provided for fixed pool sizes and tests with known sensitivity and specificity, one producing asymptotic confidence intervals (Method 3) and the other exact

confidence intervals (Method 4). These methods assume that the true values of both sensitivity and specificity are known exactly (i.e. that there is no uncertainty about the values). They do not allow for additional uncertainty in the prevalence estimate associated with uncertainty about test performance. They should be used if the test is well characterised and you can be confident that the sensitivity and specificity of the test are both close to the estimated values. If the true prevalence is likely to be close to zero, Method 4 (with exact confidence limits) should be used in preference to Method 3, because Method 3 could produce a negative lower confidence limit. Asymptotic confidence intervals (Method 3) can also be substantially narrower than exact intervals (Method 4) in some instances.

For this analysis, it was assumed that samples from 100 individual fruit bats were aggregated into 20 pools of 5 samples each, that 10 pools produced a positive test result and that the test sensitivity was 90% and specificity was 100%. An assumed sensitivity of less than 100% was used to demonstrate the possible effect of dilution on sensitivity of the pooled test. Input values and results for this analysis are summarised in the table below.

	Method	
	3	4
Input values:		
Number of pools tested	20	20
Number of pools positive	10	10
Pool size	5	5
Sensitivity	0.9	0.9
Specificity	1.0	1.0
Lower CL	0.025	0.025
Upper CL	0.975	0.975
Results:		
Est. Prevalence	0.1497	0.1497
2.5 percentile	0.0943	0.0694
97.5 percentile	0.2052	0.2818
Std. Error	0.0283	0.0283

10.4. Fixed pool size and tests with uncertain sensitivity and specificity

This method (Method 5) is for fixed pool sizes and tests with uncertain sensitivity and specificity and produces asymptotic confidence intervals about the estimated prevalence. This method assumes that the true values of both sensitivity and specificity are not known exactly and have been estimated in a limited number of samples. The analysis allows for the additional uncertainty in the prevalence estimate associated with uncertainty about test performance, based on the sample sizes used to estimate sensitivity and specificity values. This method should be used if you are uncertain about the true values of sensitivity and specificity. If the true prevalence is likely to be close to zero, the lower confidence limit could be negative.

For this analysis, it was assumed that samples from 300 individual fruit bats were aggregated into 60 pools of 5 samples each, that 29 pools produced a positive test result and that the test sensitivity was 90% and specificity was 100%. An assumed sensitivity of less than 100% was

used to demonstrate the possible effect of dilution on sensitivity of the pooled test. To allow for uncertainty about the true values of test sensitivity and specificity, it was assumed that sample sizes of 50 and 10,000, respectively, were used to estimate these values. Input values and results for this analysis are summarised in the table below.

	Method 5
Input values:	
Number of pools tested	60
Number of pools positive	29
Pool size	5
Sensitivity	0.9
Specificity	1.0
Sample size for sensitivity estimate	50
Sample size for specificity estimate	10000
Lower CL	0.025
Upper CL	0.975
Results:	
Est. Prevalence	0.1427
2.5 percentile	0.0876
97.5 percentile	0.1979
Std. Error	0.0282

10.5. Variable pool size and perfect test

This method estimates prevalence and confidence limits for variable pool sizes and assumed 100% test sensitivity and specificity. This method should be used if you can be confident that the sensitivity and specificity of the test are both close to 100%, and if pool sizes vary.

For this analysis, it was assumed that samples from 210 individual fruit bats were aggregated into 40 pools of 5 samples each with 10 samples tested individually, that 20 pools and 2 individual samples produced a positive test result and that the test sensitivity and specificity were both 100%. Input values and results for this analysis are summarised in the table below.

Variable pool size		
Input values:		
Confidence level:	0.95	
Pool size	Number of pools tested	Number of pools positive
5	40	20
1	10	2
Results:		
Est. Prevalence	0.1339	
2.5 percentile	0.0871	
97.5 percentile	0.1930	

10.6. Pooled prevalence using a Gibbs sampler

This method estimates prevalence for a fixed pool size and tests with uncertain sensitivity and specificity, using a Bayesian approach and a Gibbs sampler. It assumes that the true values of both sensitivity and specificity are not known exactly but can be estimated as Beta probability distributions. This method should be used if you are uncertain about the true values of sensitivity and specificity but can estimate their values from existing data or expert opinion. It is also useful if you already have some information on probable prevalence, which can also be included in the analysis as a prior probability distribution. This method also produces revised estimates of sensitivity and specificity, consistent with the observed data.

For this analysis, input values similar to those for the frequentist method for fixed pool size and uncertain sensitivity and specificity were used, so that results from the two approaches can be compared. It was assumed that samples from 300 individual fruit bats were aggregated into 60 pools of 5 samples each, that 29 pools produced a positive test result and that the test sensitivity was 90% and specificity was 100%. An assumed sensitivity of less than 100% was used to demonstrate the possible effect of dilution on sensitivity of the pooled test. To allow for uncertainty about the true values of test sensitivity and specificity, alpha and beta values for the prior distributions were calculated assuming that sample sizes of 50 and 10,000, respectively, were used to estimate these values. A uniform prior distribution (all values between 0 and 1 occur with equal probability) was assumed for prevalence because there was no prior information on which to base an estimate. Input values and results for this analysis are summarised in the tables below.

Input	Value
Pool size	5
Number of pools tested	60
Number of pools positive	29
Prior prevalence alpha	1
Prior prevalence beta	1
Prior Se alpha	46
Prior Se beta	6
Prior Sp alpha	10001
Prior Sp beta	1
Iterations	25000
Discard	5000

The prior Beta distributions defined above are equivalent to:

Distribution	Alpha value	Beta value	2.5% percentile	Median	97.5% percentile	Mean	Mode	Standard deviation
Prevalence	1	1	0.025	0.5	0.975	0.5		0.2887
Sensitivity	46	6	0.7859	0.8895	0.9556	0.8846	0.9	0.0439
Specificity	10001	1	0.9996	0.9999	1	0.9999	1	0.0004

The simulation was run for 25,000 iterations, with 5,000 iterations discarded to allow for convergence. Posterior probability distributions for prevalence, sensitivity and specificity are summarised below. Median and upper and lower 95% probability limits from this analysis were all slightly higher than the corresponding values from the frequentist approach.

Summary	Prevalence	Sensitivity	Specificity

results			
Minimum	0.0651	0.6492	0.999
0.025	0.1015	0.7775	0.9996
Median	0.1509	0.8847	0.9999
0.975	0.226	0.9536	1
Maximum	0.3267	0.9951	1
Mean	0.1533	0.8808	0.9999
SD	0.032	0.0454	0.0004
Iterations	20000	20000	20000

10.7. Estimated true prevalence using one test (unpooled) with a Gibbs sampler

This method uses a Bayesian approach and Gibbs sampling to estimate the true animal-level prevalence of infection based on testing of individual (not pooled) samples using a test with imperfect sensitivity and/or specificity. As for the Bayesian method for pooled sampling, the analysis requires prior estimates of true prevalence, test sensitivity and test specificity as Beta probability distributions, and outputs posterior distributions for prevalence, sensitivity and specificity. This method is preferable to the conventional (Rogan-Gladen) method for estimating true prevalence, because it allows for uncertainty about the true values for sensitivity and specificity when calculating probability limits for the true prevalence estimate, which are not routinely included in the conventional approach. It also allows incorporation of prior information on the likely true prevalence based on pre-existing estimates or expert opinion.

For this analysis, the original values for stool sampling for *Strongyloides* infection in Cambodian refugees from Joseph et al. (1996) were used, as listed in the table below, and 95% probability limits were calculated about the estimated prevalence.

Input	Value
Number tested	162
Number test +ve	40
Prior prevalence alpha	1
Prior prevalence beta	1
Prior Se alpha	4.44
Prior Se beta	13.31
Prior Sp alpha	71.25
Prior Sp beta	3.75
Iterations	25000
Discard	5000
True pos start	35
False neg start	35

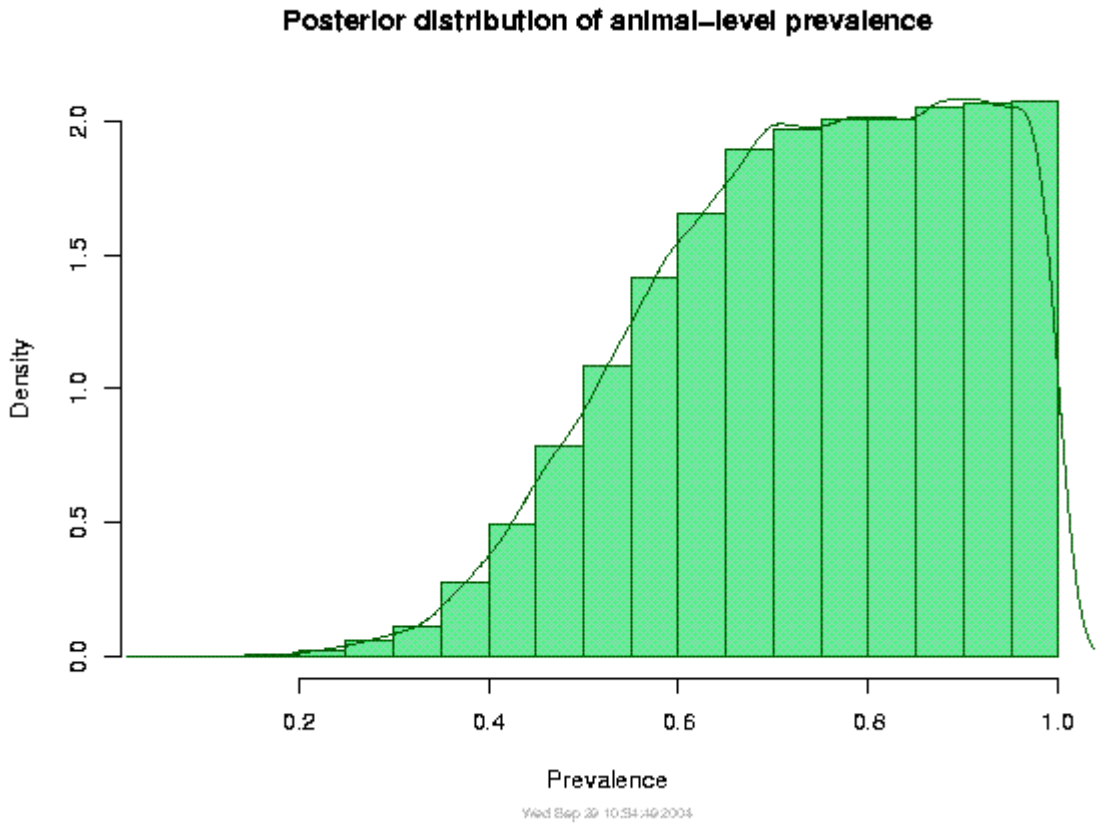
The prior Beta distributions defined above are equivalent to:

Distribution	Median	95% upper limit	95% lower limit
Prevalence	0.5	0.025	0.975
Sensitivity	0.24	0.07	0.47
Specificity	0.95	0.89	0.99

Distribution	Alpha value	Beta value	2.5% percentile	Median	97.5% percentile	Mean	Mode	Standard deviation
Prevalence	1	1	0.025	0.5	0.975	0.5		0.2887
Sensitivity	4.44	13.31	0.0843	0.2406	0.469	0.2501	0.2184	0.1
Specificity	71.25	3.75	0.8909	0.954	0.9868	0.95	0.9623	0.025

The simulation was run for 25,000 iterations, with 5,000 iterations discarded to allow for convergence. Posterior probability distributions for prevalence, sensitivity, specificity and other parameters from the analysis are summarised below.

	Prevalence	Sensitivity	Specificity	PPV	NPV	LR for positive	LR for negative	True positives	False negatives
Minimum	0.171	0.135	0.8	0.197	0.243	1	0.32	7	7
0.025	0.393	0.212	0.882	0.665	0.336	2.4	0.54	29	33
Median	0.738	0.307	0.951	0.883	0.538	6.4	0.73	38	82
0.975	0.985	0.484	0.986	0.969	0.786	24.4	0.85	40	120
Maximum	1	0.697	0.998	0.994	0.907	157.8	1	40	122
Mean	0.728	0.316	0.948	0.871	0.544	7.5	0.72	38	81
SD	0.165	0.07	0.027	0.08	0.124	6.4	0.08	3	25
Iterations	20000	20000	20000	20000	20000	20000	20000	20000	20000



10.8. Estimated true prevalence using two tests (unpooled) with a Gibbs sampler

This method uses a Bayesian approach and Gibbs sampling to estimate the true animal-level prevalence of infection based on testing of individual (not pooled) samples using two independent tests with imperfect sensitivity and/or specificity. The analysis requires prior estimates of true prevalence and test sensitivity and test specificity for both tests, as Beta probability distributions. It outputs posterior distributions for prevalence, sensitivity and specificity of both tests and several other parameters of interest. This method is similar to the one-test method, except that it allows incorporation of data from two tests used concurrently, and finds the best estimate that fits the combination of the prior information and the observed data. It also allows for uncertainty about the true values for sensitivity and specificity when calculating probability limits for the true prevalence estimate and the incorporation of prior information on the likely true prevalence based on pre-existing estimates or expert opinion. Because of the use of two tests, this method will often produce narrower probability limits about the prevalence estimate than the one-test method, particularly where there is considerable uncertainty about prior estimates.

For this analysis, the original values for stool sampling and serology for *Strongyloides* infection in Cambodian refugees from Joseph et al. (1996) were used, as listed in the table below, and 95% probability limits were calculated about the estimated prevalence.

Input	Value
a (T1+/T2+)	38
b (T1+/T2-)	87
c (T1-/T2+)	2
d (T1-/T2-)	35
P alpha	1
P beta	1
Se 1 alpha	21.96
Se 1 beta	5.49
Sp 1 alpha	4.1
Sp 1 beta	1.76
Se 2 alpha	4.44
Se 2 beta	13.31
Sp 2 alpha	71.25
Sp 2 beta	3.75
Y1 start	35
Y2 start	30
Y3 start	2
Y4 start	10
Iterations	25000
Discard	5000

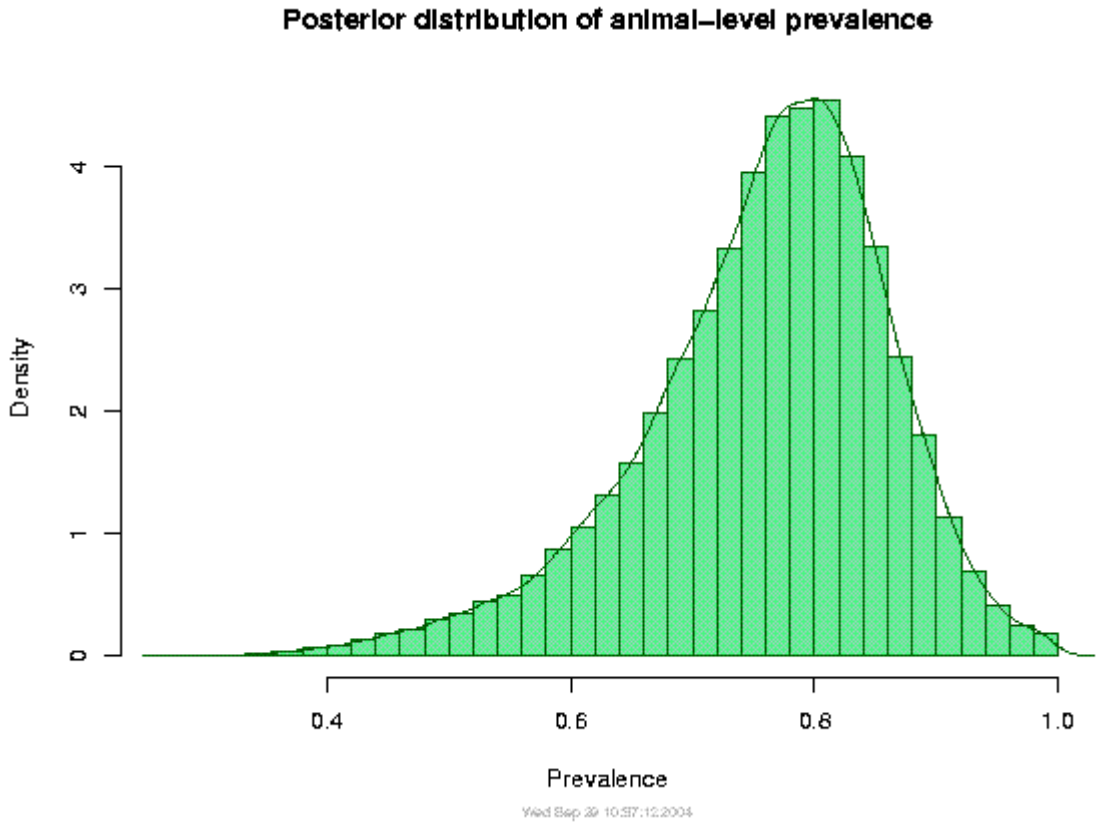
The prior Beta distributions defined above are equivalent to:

Distribution	Alpha value	Beta value	2.5% percentile	Median	97.5% percentile	Mean	Mode	Standard deviation
Prevalence	1	1	0.025	0.5	0.975	0.5		0.2887
Sensitivity (serology)	21.96	5.49	0.6346	0.8073	0.9242	0.8	0.8236	0.075

Specificity (serology)	4.1	1.76	0.3123	0.7235	0.9621	0.6997	0.8031	0.175
Sensitivity (stool)	4.44	13.31	0.0843	0.2406	0.469	0.2501	0.2184	0.1
Specificity (stool)	71.25	3.75	0.8909	0.954	0.9868	0.95	0.9623	0.025

The simulation was run for 25,000 iterations, with 5,000 iterations discarded to allow for convergence. Posterior probability distributions for prevalence, sensitivity, specificity and other parameters from the analysis are summarised below.

	Prevalence	Test 1 Se	Test 1 Sp	Test 1 PPV	Test 1 NPV	Test 2 Se	Test 2 Sp	Test 2 PPV	Test 2 NPV	Y1	Y2	Y3	Y4
Minimum	0.275	0.689	0.129	0.402	0.021	0.135	0.823	0.565	0.292	26	16	0	0
0.025	0.528	0.791	0.376	0.679	0.277	0.222	0.906	0.775	0.389	34	45	0	2
Median	0.772	0.889	0.695	0.918	0.639	0.305	0.96	0.907	0.519	38	77	2	10
0.975	0.921	0.954	0.955	0.992	0.858	0.425	0.989	0.973	0.705	38	87	2	25
Maximum	0.998	0.984	0.999	1	0.959	0.615	0.998	0.997	0.895	38	87	2	35
Mean	0.761	0.885	0.688	0.9	0.624	0.308	0.958	0.9	0.525	37	74	2	10
SD	0.099	0.042	0.16	0.084	0.15	0.051	0.021	0.051	0.079	1	11	1	6
Iterations	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000



10.9. Sample size calculation for fixed pool size and perfect tests

This option can be used to calculate the number of pools that need to be tested for various values for pool size, estimated prevalence and desired confidence and precision of the estimate. It assumes fixed pool sizes and a test with 100% sensitivity and specificity.

For this analysis, sample size was calculated to provide 95% confidence of estimating a true prevalence of 0.14 (14%) with a precision of 0.055 (5.5%). This is equivalent to the observed prevalence and precision when 162 samples from little red flying foxes in Queensland were tested individually, with 22 positive results (H. Field, pers com). Input values and summary results are shown below. For a pool size of five, 42 pools would be sufficient to provide the desired precision and confidence in the estimated prevalence.

Input	Value
Prevalence	0.14
Precision	0.055
Confidence	0.95

Pool size	Number of pools	Total number of individual samples
1	153	153
2	83	166
5	42	210
10	33	330
15	36	540
20	46	920
25	64	1600
30	95	2850

10.10. Sample size calculation for fixed pool size and tests with known sensitivity and specificity

This option can be used to calculate the number of pools that need to be tested for various values for pool size, estimated prevalence and desired confidence and precision of the estimate. It assumes fixed pool sizes and that the true values of both sensitivity and specificity are known exactly (i.e. that there is no uncertainty about the values).

For this analysis, sample size was calculated to provide 95% confidence of estimating a true prevalence of 0.14 (14%) with a precision of 0.055 (5.5%). This is equivalent to the observed prevalence and precision when 162 samples from little red flying foxes in Queensland were tested individually, with 22 positive results (H. Field, pers com). Test sensitivity was assumed to be 90%, to demonstrate the effect of imperfect test sensitivity on sample size estimates and specificity was assumed to be 100%. For a pool size of five, 20 pools would be sufficient to provide the desired precision and confidence in the estimated prevalence, assuming asymptotic confidence limits were used. Suggested sample sizes are generally lower than those for a perfect test with corresponding pool sizes and the resulting asymptotic confidence limits van be substantially narrower than corresponding exact limits.

Input	Value
Prevalence	0.14
Precision	0.055
Confidence	0.95
Sensitivity	0.9
Specificity	1

Pool size	Number of pools	Total number of individual samples
1	173	173
2	82	164
5	20	100
10	5	50

10.11. Sample size calculation for fixed pool size and tests with uncertain sensitivity and specificity

This option can be used to calculate the number of pools that need to be tested for various values for pool size, estimated prevalence and desired confidence and precision of the estimate. It assumes fixed pool sizes and that the true values of both sensitivity and specificity are not known exactly and have been estimated in a limited number of samples.

For this analysis, sample size was calculated to provide 95% confidence of estimating a true prevalence of 0.14 (14%) with a precision of 0.055 (5.5%). This is equivalent to the observed prevalence and precision when 162 samples from little red flying foxes in Queensland were tested individually, with 22 positive results (H. Field, pers com). Test sensitivity was assumed to be 90%, to demonstrate the effect of imperfect test sensitivity on sample size estimates and specificity was assumed to be 100%. To allow for uncertainty about the true values of test sensitivity and specificity, it was assumed that sample sizes of 50 and 10,000, respectively, were used to estimate these values. For a pool size of five, 59 pools would be sufficient to provide the desired precision and confidence in the estimated prevalence.

Input	Value
Prevalence	0.14
Precision	0.055
Confidence	0.95
Sensitivity	0.9
Specificity	1
Sensitivity sample size	50
Specificity sample size	10000

Pool size	Number of pools	Total number of individual samples
1	183	183
2	102	204
5	59	295
10	67	670

15	238	3570

10.12. Simulate sampling for fixed pool size and assumed perfect test

This program simulates sampling and prevalence estimation for a specified (design) prevalence value and level of confidence. The program runs multiple iterations of sampling, pooling and testing from an infinite population with the specified prevalence, estimates true prevalence assuming a prefect test (using Method 2) for each iteration and calculates the mean prevalence and estimated bias across all iterations. It assumes fixed pool sizes and a test with 100% sensitivity and specificity. Values for the true sensitivity and specificity that are different to the assumed values of 100% can also be entered if desired, to check the importance of the assumption of a perfect test

For this analysis, six alternative pooling strategies were evaluated for the estimation of prevalence in a population with an assumed true prevalence of 0.14 (14%). Pool sizes and numbers of pools were previously estimated to provide 95% confidence of estimating a true prevalence of 0.14 with a precision of 0.055 (see sample size examples). This is equivalent to the observed prevalence and precision when 162 samples from little red flying foxes in Queensland were tested individually, with 22 positive results (H. Field, pers com). The true sensitivity and specificity of the test were both assumed to be 1 (100%), equal to the assumed values for prevalence estimation. Input values, pooling strategies and results are summarised in the tables below:

Input	Value
Method	Fixed pool size and known Se & Sp
Assumed Prevalence	0.14
Assumed Sensitivity	1
Assumed Specificity	1
True Sensitivity	1
True Specificity	1
Confidence	0.95
No. of strategies	6
No. of iterations	1000

Strategy	Pool size	Number of pools
1	2	83
2	3	60
3	4	49
4	5	42
5	10	33
6	20	46

Strategy	Mean prevalence	Minimum prevalence	Maximum prevalence	Mean bias	Mean CI width	Mean Std. Error	Mean Sq. Error	Bias/AP	Bias/TP	Bias/MSE	Proportion Valid
1	0.14101	0.06862	0.23165	0.00101	0.11593	0.02794	0.00079	0.00719	0.00724	0.0013	0.968
2	0.14017	0.06528	0.22434	0.00017	0.1161	0.02787	0.00079	0.00124	0.00124	4e-05	0.96
3	0.14037	0.06782	0.25617	0.00037	0.11666	0.02789	0.00079	0.0026	0.00261	0.00017	0.967

4	0.14123	0.06508	0.23506	0.00123	0.11878	0.02826	0.00081	0.00872	0.0088	0.00187	0.965
5	0.14534	0.06983	1	0.00534	0.12772	0.02923	0.00092	0.03671	0.03811	0.03097	0.972
6	0.23773	0.08374	1	0.09773	0.22332	0.02911	0.01046	0.41109	0.69804	0.91322	0.974

10.13. Simulate sampling for fixed pool size and test with known sensitivity and specificity

This program simulates sampling and prevalence estimation for a specified (design) prevalence value and level of confidence. The program runs multiple iterations of sampling, pooling and testing from an infinite population with the specified prevalence, estimates true prevalence assuming known test sensitivity and specificity (using Method 4) for each iteration and calculates the mean prevalence and estimated bias across all iterations. It assumes fixed pool sizes and that the true values of both sensitivity and specificity are known exactly (i.e. that there is no uncertainty about the values). Values for the true sensitivity and specificity that are different to the assumed values can also be entered if desired to check the importance of the assumption of a perfect test

For this analysis, five alternative pooling strategies were evaluated for the estimation of prevalence in a population with an assumed true prevalence of 0.14 (14%). Pool sizes and numbers of pools were previously estimated to provide 95% confidence of estimating a true prevalence of 0.14 with a precision of 0.055 (see sample size examples). This is equivalent to the observed prevalence and precision when 162 samples from little red flying foxes in Queensland were tested individually, with 22 positive results (H. Field, pers com). The sensitivity and specificity of the test were assumed to be 0.9 (90%) and 1 (100%) for prevalence estimation, and the true values were assumed to be the same as the assumed values for prevalence estimation. Input values, pooling strategies and results are summarised in the tables below:

Input	Value
Method	Fixed pool size and known Se & Sp
Assumed Prevalence	0.14
Assumed Sensitivity	0.9
Assumed Specificity	1
True Sensitivity	0.9
True Specificity	1
Confidence	0.95
No. of strategies	5
No. of iterations	1000

Strategy	Pool size	Number of pools
1	2	82
2	3	47
3	4	30
4	5	20
5	10	5

Mean confidence interval widths are greater than the target value of 0.11 (± 0.55) because exact methods were used to calculate confidence limits in these simulations, rather than the asymptotic methods used ti estimate sample size.

Strategy	Mean prevalence	Minimum prevalence	Maximum prevalence	Mean bias	Mean CI width	Mean Std. Error	Mean Sq. Error	Bias/AP	Bias/TP	Bias/MSE	Proportion Valid
1	0.14102	0.05575	0.28432	0.00102	0.12572	0.02788	0.00078	0.0072	0.00725	0.00131	0.964
2	0.14285	0.03257	0.2875	0.00285	0.14582	0.02787	0.00079	0.01995	0.02036	0.01028	0.967
3	0.14153	0.04991	0.26221	0.00153	0.17028	0.02749	0.00076	0.01082	0.01094	0.00308	0.978
4	0.14581	0.03581	0.35561	0.00581	0.21857	0.0275	0.00079	0.03983	0.04149	0.04256	0.98
5	0.1377	0	0.19726	-0.0023	0.83581	0.0252	0.00064	-0.01672	-0.01644	0.00824	0.821

10.14. Simulate sampling for fixed pool size and test with uncertain sensitivity and specificity

This program simulates sampling and prevalence estimation for a specified (design) prevalence value and level of confidence. The program runs multiple iterations of sampling, pooling and testing from an infinite population with the specified prevalence, estimates true prevalence assuming uncertain test sensitivity and specificity (using Method 5) for each iteration and calculates the mean prevalence and estimated bias across all iterations. It assumes fixed pool sizes and that the true values of both sensitivity and specificity are not known exactly and have been estimated in a limited number of samples. Values for the true sensitivity and specificity that are different to the assumed values can also be entered if desired to check the importance of the assumption of a perfect test

For this analysis, six alternative pooling strategies were evaluated for the estimation of prevalence in a population with an assumed true prevalence of 0.14 (14%). Pool sizes and numbers of pools were previously estimated to provide 95% confidence of estimating a true prevalence of 0.14 with a precision of 0.055 (see sample size examples). This is equivalent to the observed prevalence and precision when 162 samples from little red flying foxes in Queensland were tested individually, with 22 positive results (H. Field, pers com). The sensitivity and specificity of the test were assumed to be 0.9 (90%) and 1 (100%) for prevalence estimation, and the true values were assumed to be the same as the assumed values for prevalence estimation. To allow for uncertainty about the true values of test sensitivity and specificity, it was assumed that sample sizes of 50 and 10,000, respectively, were used to estimate these values. Input values, pooling strategies and results are summarised in the tables below:

Input	Value
Method	Uncertain test Se & Sp
Assumed Prevalence	0.14
Assumed Sensitivity	0.9
Assumed Specificity	1
Sample size for sensitivity	50
Sample size for specificity	10000
True Sensitivity	0.9
True Specificity	1
Confidence	0.95
No. of strategies	6
No. of iterations	1000

Strategy	Pool size	Number of pools
1	2	102
2	3	76
3	4	65
4	5	59
5	10	67
6	15	238

Strategy	Mean prevalence	Minimum prevalence	Maximum prevalence	Mean bias	Mean CI width	Mean Std. Error	Mean Sq. Error	Bias/AP	Bias/TP	Bias/MSE	Proportion Valid
1	0.13979	0.06765	0.2416	-0.00021	0.10931	0.02789	0.00079	-0.00149	-0.00149	6e-05	0.956
2	0.14149	0.06785	0.25397	0.00149	0.11044	0.02817	0.00081	0.01054	0.01065	0.00276	0.954
3	0.14229	0.05578	0.24016	0.00229	0.11073	0.02825	0.00082	0.01613	0.01639	0.00644	0.961
4	0.14266	0.07427	0.26878	0.00266	0.11168	0.02849	0.00084	0.01861	0.01896	0.00841	0.958
5	0.14404	0.07286	0.27866	0.00404	0.11897	0.03035	0.00103	0.02805	0.02886	0.01584	0.978
6	0.14229	0.10554	0.37192	0.00229	0.12771	0.03258	0.00642	0.01606	0.01632	0.00081	0.998

10.15. Simulate sampling for variable pool size and assumed perfect test

This program simulates sampling and prevalence estimation for a specified (design) prevalence value and level of confidence. The program runs multiple iterations of sampling, pooling and testing from an infinite population with the specified prevalence, estimates true prevalence for each iteration and calculates the mean prevalence and estimated bias across all iterations. It assumes variable pool sizes and a test with 100% sensitivity and specificity. Values for the true sensitivity and specificity that are different to the assumed values of 100% can also be entered if desired to check the importance of the assumption of a perfect test

For this analysis, six alternative pooling strategies were evaluated for the estimation of prevalence in a population with an assumed true prevalence of 0.14 (14%). This is equivalent to the observed prevalence and precision when 162 samples from little red flying foxes in Queensland were tested individually, with 22 positive results. Pool sizes and numbers of pools were used to provide the same total sample size (210 samples) as used for the fixed pool size and perfect test example. The true sensitivity and specificity of the test were both assumed to be 1 (100%), equal to the assumed values for prevalence estimation. Input values, pooling strategies and results are summarised in the tables below:

Input	Value
Method	Variable pool size & perfect test
Assumed Prevalence	0.14
Assumed Sensitivity	1
Assumed Specificity	1
True Sensitivity	1
True Specificity	1
Confidence	0.95
No. of strategies	6
No. of iterations	1000

Strategy	Pool size 1	Number of pools 1	Pool size 2	Number of pools 2
1	5	42	0	0
2	5	40	1	10
3	5	40	10	1
4	10	21	0	0
5	10	20	1	10
6	20	10	10	1

Strategy	Mean prevalence	Minimum prevalence	Maximum prevalence	Mean bias	Mean CI width	Mean Std. Error	Mean Sq. Error	Bias/AP	Bias/TP	Bias/MSE	Proportion Valid
1	0.14272	0.05293	0.28226	0.00272	0.11136	NaN	NaN	0.01909	0.01946	NaN	0.945
2	0.14256	0.05826	0.27087	0.00256	0.11016	NaN	NaN	0.01793	0.01826	NaN	0.935
3	0.14161	0.05986	0.25929	0.00161	0.11187	NaN	NaN	0.01139	0.01153	NaN	0.97
4	0.15136	0.06262	0.92587	0.01136	0.14941	NaN	NaN	0.07507	0.08116	NaN	0.935
5	0.14654	0.06912	0.32703	0.00654	0.1433	NaN	NaN	0.0446	0.04669	NaN	0.949
6	0.49031	0.04148	0.89974	0.35031	0.18629	NaN	NaN	0.71447	2.50225	NaN	0.49

The standard error of the estimate cannot be calculated using this method, so that and other measures derived from it (Mean sq. error and Bias/MSE) are listed as 'NaN' (Not a number).

10.16. Demonstration of freedom using pooled testing with tests of known sensitivity and fixed pool size

This program estimates the confidence that a population is free of disease, or of detecting disease at a specified design prevalence, using pooled testing and assuming a fixed pool size, a test of known sensitivity and 100% specificity and that all pools have a negative test result.

For this analysis, it was assumed that 30 pools of 10 were tested using a test with a sensitivity of 0.9 (90%) and perfect specificity. The design prevalence which we wish to detect is 0.01 (1%) and we require 95% confidence of detecting one or more positives if the true prevalence is greater than or equal to the design prevalence. Input values and results are summarised below:

Input values

Pool size	10
Number of pools	30
Sensitivity	0.9
Confidence level	0.95
Design prevalence	0.01

Results

If all pools test negative, there is **95%** confidence that the true prevalence is less than **0.0111**.

If all pools test negative, there is **93.28%** confidence that the true prevalence is less than **0.01**.

For a pool size of **10**, a minimum of **34** pools must be tested to provide **95%** confidence of detecting a prevalence of **0.01**.

The table below lists the number of pools required for various pool sizes to provide **95%** confidence of detecting a prevalence of **0.01** and assuming a test sensitivity of **0.9** for all pool sizes.

Pool size	Number of pools
1	332
2	166
3	111
4	83
5	67
10	34
15	23
20	17
25	14
30	12
40	9
50	7
100	4

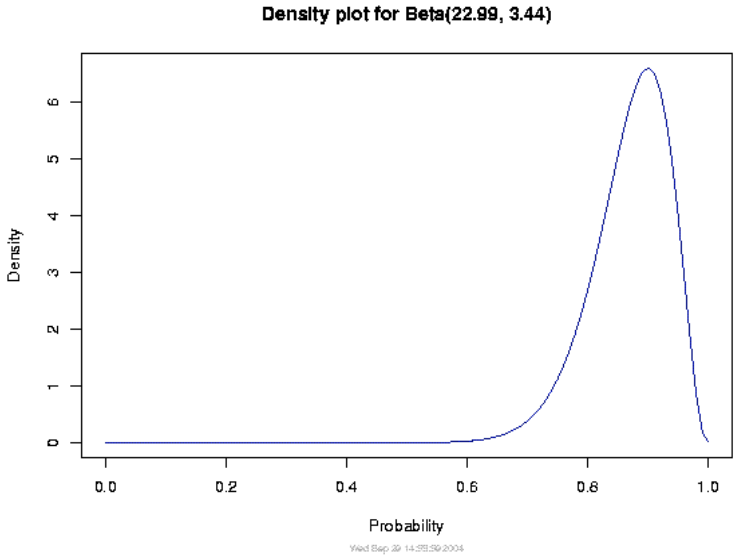
10.17. Estimation of alpha and beta Parameters for Prior Beta distributions

This program calculates the alpha and beta parameters for prior Beta distributions used in the Bayesian analyses, based on the values specified for the mode and 5th or 95th percentile of the distribution. The mode and percentiles can be estimated either from existing data or from expert opinion.

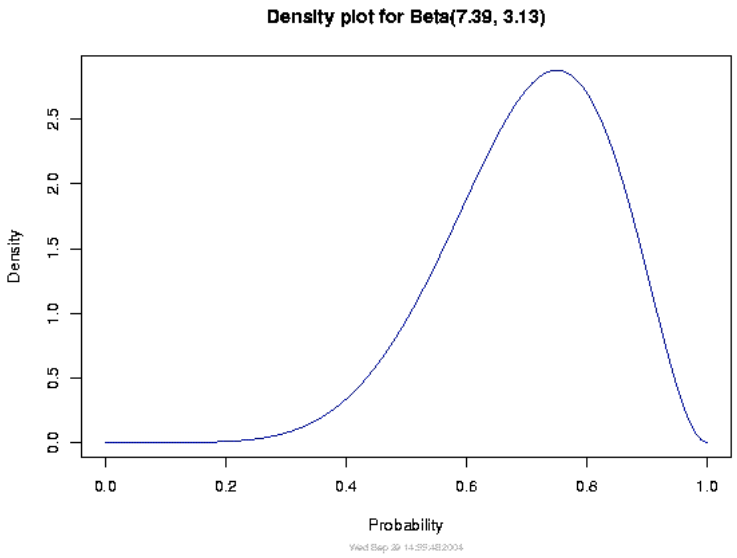
For this analysis, two estimates were calculated. The first for a distribution with a mode of 0.9 (90%) and a 5th percentile of 0.75 (75%), and the second is for a distribution with a mode at 0.75 and the 95th percentile at 0.9. input values and results are summarised below:

Distribution	Mode	5/95 percentile	Alpha value	Beta value
1	0.9	0.75	22.99	3.44
2	0.75	0.9	7.39	3.13

Density curves for the two distributions are:



and:



10.18. Estimation of Beta probability distributions for specified alpha and beta parameters

This utility calculates summary values for Beta distributions for specified values for alpha and beta parameters.

For this analysis, two estimates were calculated. The first is for a distribution with alpha = 22.99 and beta = 3.44, and the second for a distribution with alpha = 7.39 and beta = 3.13. These values correspond to the parameters estimated in the parameter estimation example. Input values and results are summarised below:

Alpha value	Beta value	2.5% percentile	25% percentile	Median	75% percentile	97.5% percentile	Mean	Mode	Standard deviation
22.99	3.44	0.7207	0.832	0.8792	0.9176	0.9668	0.8698	0.9001	0.0642
7.39	3.13	0.41	0.6144	0.7157	0.804	0.9228	0.7025	0.75	0.1347

11. Appendix 2: Case studies – Hendra/Nipah virus

11.1. Introduction

Three case studies were undertaken to evaluate the potential role of pooled prevalence estimation in cases of emerging diseases. These case studies were used to demonstrate the potential application of pooled testing under scenarios of high, moderate and low prevalence of disease.

11.2. Methods

Pooled prevalence calculation

For these case studies, simulations and calculation of prevalence estimates and exact confidence limits were undertaken using Methods 2 and 4 from the Pooled Prevalence Calculator (http://www.ausvet.com.au/pprev). These methods are equivalent to Methods 3 and 5 from Cowling et al. 1999, and assume that:

- pool size is constant for each pooling strategy evaluated;
- test sensitivity and specificity are known exactly and are either 1 (100%) (Method 2) or < 1 (Method 4);
- individual samples were selected and allocated to pools on a random basis;
- the outcome of interest follows a binomial distribution in the population; and
- sample size is small relative to the population of interest.

Numbers of pools required to be tested for various pool sizes and scenarios were calculated using the sample size calculation option from the Pooled Prevalence Calculator. These methods calculate sample sizes for perfect or imperfect tests (with known sensitivity and specificity) by transposition of formulae for calculation of asymptotic confidence limits about pooled prevalence estimates (Methods 1 and 3 for pooled prevalence estimation).

Summary of case study data

Three case studies were undertaken to evaluate the potential role of pooled prevalence estimation in cases of emerging diseases. The first two case studies relate to high and moderate-prevalence situations with Hendra virus in fruit bats in Australia, whereas the third case study relates to a low-prevalence situation, with infection of abattoir workers in Malaysia with Nipah virus from infected pigs (see Table 1). In all three cases, individual samples were tested to determine a test result for each sample, and hence estimate prevalence. Sampling was opportunistic for the first two case studies and by voluntary participation for the third. However, for the pooled testing evaluation, samples were assumed to be representative of their respective populations.

Simulated pooling strategies

For each case study, pooled testing was simulated as follows:

1. Up to six different pool sizes were calculated that each divided equally (or nearly equally) into the total sample size;
2. A population the size of the original sample was simulated to contain the same number of positive individuals as in the original individual test results;

3. For each pool size, the appropriate number of pools was calculated so that the whole sample population would be tested and individuals for each pool were selected from the population at random, without replacement, until the population had all been allocated to pools;
4. For pool sizes that did not divide equally into the population size, sampling and allocation to pools continued until there were too few remaining individuals to form another pool;
5. For each pool, the test result was determined based on whether or not positive individuals had been included in the pool, and the assumed values used for test sensitivity and specificity.
6. For each pool size, pooled prevalence and exact confidence limits were estimated based on pool size, the number of pools tested, the number of pools positive and the assumed values of test sensitivity and specificity;
7. For each pool size, sampling and testing were simulated for 1,000 iterations and the mean prevalence estimate, mean bias and mean confidence interval width were calculated; and
8. Simulations were repeated assuming known sensitivities (fixed values) of both 1 and 0.9 (allowing for a reduction in sensitivity associated with pooling). Specificity was set at 1 for all simulations.

Table 1: Summary of three case studies for evaluating pooled testing for estimating prevalence

Case study	Host population	Time period	Number tested	Number positive	Estimated prevalence (%)	Binomial confidence limits (%)	Source
1	Black flying fox (*Pteropus alecto*) in Queensland	April 1996 – August 1999	585	343	58.6	55 – 63	H. Field (pers com)
2	Little red flying fox (*Pteropus scapulatus*) in Queensland	April 1996 – August 1999	162	22	13.6	8.7 – 19.8	H. Field (pers com)
3	Human abattoir workers killing pigs in Malaysia	1998 – 1999	435	7	1.6	0.6 – 3.3	Sahani M et al., 2001

Sample size calculations

For each case study, sample sizes (pool size and number of pools) for a range of pool sizes were estimated assuming a perfect test (sensitivity and specificity both = 1) to provide 95% confidence of estimating prevalence with an assumed prevalence and precision similar to those observed with the individual testing.

Simulation of population sampling

Finally, sampling from an infinite population was simulated for each case study using sample sizes and prevalence values from the sample size calculations. These simulations were similar to those for evaluating pooling strategies (above), except that they assumed sampling from an infinite population with the specified prevalence, rather than from a finite population with a fixed number of positive individuals. Again, simulations assuming both 1.0 and 0.9 sensitivity were run to evaluate the possible effect of reduced sensitivity of the test due to dilution in pooled samples. For each simulation, the mean bias, mean 95% confidence interval width and the proportion of estimates where the 95% confidence interval contained the assumed true

(design) prevalence were calculated. Also for each case study, the most frequent number of positive results was determined from the simulation for a selected pool size, and the prevalence and confidence limits estimated for this result.

11.3. Results

Simulated pooling strategies
The results of simulated pooling strategies are summarised in Table 2. For all three case studies, mean prevalence estimates were very similar to the original unpooled estimates for small pool sizes for both perfect tests and assumed sensitivity of 0.9. For the high and moderate-prevalence cases studies (Case Studies 1 and 2) the mean prevalence was upwardly biased at larger pool sizes, except for the simulations assuming imperfect sensitivity, where estimated prevalence was negatively biased at larger pool sizes and higher prevalence. The negative bias in these simulations was due to the exclusion of many iterations in which more pools tested positively than was consistent with the estimated sensitivity of the test, so that prevalence could not be estimated for these iterations. The effect of pool size on bias in the prevalence estimate was most apparent in the high prevalence situation (Case Study 1), where there was substantial bias at a pool size of 5, and least apparent in the low-prevalence situation (Case Study 3), where there was negligible bias even at a pool size of 29.

Mean width of confidence intervals was similar to the original unpooled estimates for small pool sizes and increased as pool size increased. The effect of pool size on confidence interval width was also related to prevalence, with a greater effect at higher prevalence.

Sample size calculations
The estimated numbers of pools to be tested to provide the desired confidence and precision for prevalence estimates for each of the three case studies are summarised in Table 3. Generally, the number of pools required to be tested decreased as pool size increased, except for the high-prevalence situation (Case Study 1). For this situation the numbers of pools increased with pool size because of the high proportion of pools that would test positively, so that more pools have to be tested to ensure that sufficient pools will test negatively to allow prevalence estimation. For some larger pool sizes, the number of pools to be tested could not be calculated or were too large (or small) to be of practical use for prevalence estimation (shown in the table as N/A).

At smaller pool sizes, the total sample sizes required were generally similar to or slightly more than was required for individual testing. As pool size increased, so did the total sample size for all situations examined. However, the total number of samples to be tested remained considerably less than the individual testing option, except for Case Study 1. The numbers of pools to be tested decreased for each pool size if sensitivity was assumed to be less than 1, except for Case Study 3, where the numbers of pools increased slightly, depending on pool size. In fact, use of an imperfect test resulted in considerably smaller numbers of pools for Case Study 1, because of the reduced probability of all pools testing positively. For Case Study 1, relaxation of the desired precision would also substantially reduce the required sample size. For example, relaxing the precision from ± 0.04 to ± 0.05 (a very small change) substantially reduces the numbers of pools to be tested to 320, 388 and 555 (from 499, 606 and 867), for pool sizes of 2, 3 and 4, respectively.

Table 2: Mean prevalence and mean 95% confidence interval width for three case studies at various pool sizes and for both perfect tests and for a test with known sensitivity of 0.9 and specificity of 1.0

			Perfect test			Known Se = 0.9, Sp = 1		
Strategy	Pool size	Number of pools	Mean prevalence	Mean bias	Mean CI width	Mean prevalence	Mean bias	Mean CI width
Case study 1								
Original sampling	1	585	0.586	-	0.08			
1	2	292	0.587	0.001	0.108	0.588	0.001	0.140
2	3	195	0.589	0.003	0.150	0.596	0.010	0.253
3	4	146	0.603	0.017	0.227	0.583	-0.003	0.311
4	5	117	0.671	0.085	0.350	0.537	-0.050	0.359
5	9	65	0.985	0.399	0.716	0.337	-0.249	0.229
Case study 2								
Original sampling	1	162	0.136	-	0.111			
1	2	81	0.136	0.000	0.116	0.136	0.000	0.125
2	3	54	0.137	0.001	0.122	0.137	0.001	0.133
3	6	27	0.140	0.004	0.143	0.139	0.003	0.167
4	9	18	0.143	0.007	0.175	0.145	0.009	0.198
5	18	9	0.562	0.426	0.589	0.152	0.016	0.141
6	27	6	0.936	0.801	0.917	0.078	-0.057	0.125
Case study 3								
Original sampling	1	435	0.016	-	0.027			
1	2	217	0.016	0.000	0.027	0.016	0.000	0.028
2	3	145	0.016	0.000	0.027	0.016	0.000	0.028
3	5	87	0.016	0.000	0.027	0.016	0.000	0.029
4	10	43	0.016	0.000	0.028	0.016	0.000	0.030
5	15	29	0.016	0.000	0.029	0.016	0.000	0.031
6	29	15	0.017	0.001	0.031	0.017	0.001	0.036

Simulation of population sampling

The results of simulated sampling from the population are summarised in Table 4. For Case Study 1 (high prevalence), pool sizes of 2, 3 and 4 only were simulated for the assumed perfect test, because of the very large numbers of pools required at larger pool sizes. At these pool sizes, and assuming a perfect test, bias was negligible, confidence interval width was close to the target of ± 0.04 and > 95% of iterations yielded a valid estimate (where the estimated 95% confidence interval contained the design prevalence value). If test sensitivity was assumed to be only 0.9 instead of 1.0, mean bias increased with pool size (in a negative direction). Mean confidence interval width increased rapidly with pool size, whereas the proportion of valid estimates decreased rapidly for pool sizes >3. the high proportion of invalid estimates was mainly because of the large number of iterations where the number of

positive pools was not consistent with the assumed sensitivity and specificity of the test, so that it was not possible to estimate the prevalence.

Table 3: Numbers of pools to be tested and total number of samples required to provide 95% confidence of estimating true prevalence with the specified precision for three case studies and for both perfect (Sensitivity = Specificity = 1) and imperfect (Sensitivity = 0.9 and Specificity = 1) tests.

	Case Study 1: prevalence = 0.59, precision = 0.04		Case Study 2: prevalence = 0.14, precision = 0.055		Case Study 3: prevalence = 0.016, precision = 0.014	
Pool size	Number of pools	Total number of individual samples	Number of pools	Total number of individual samples	Number of pools	Total number of individual samples
Sensitivity = Specificity = 1						
1	581	581	153	153	309	309
2	499	998	83	166	156	312
3	606	1818	60	180	105	315
4	867	3468	49	196	79	316
5	N/A	N/A	42	210	64	320
10	N/A	N/A	33	330	33	330
20	N/A	N/A	46	920	18	360
Sensitivity = 0.9, Specificity = 1						
1	739	739	173	173	344	344
2	341	682	82	164	171	342
3	147	441	47	141	112	336
4	78	312	30	120	83	332
5	49	245	20	100	65	325
10	14	140	5	50	29	290
20	4	80	N/A	N/A	12	240

For Case Study 2 (moderate prevalence) with a perfect test, mean bias remained negligible for pool sizes up to 10, but was substantial for a pool size of 20. Mean confidence interval width was close to the expected value for pool sizes up to 10, but also increased substantially for a pool size of 20, so that the proportion of valid estimates remained >95%, for all pool sizes. For an assumed test sensitivity of 0.9, bias remained negligible for all pool sizes, but the mean confidence interval width increased with pool size, so that the proportion of valid estimates remained >95% for pool sizes up to 5, but not for a pool size of 10.

For Case Study 3 (low prevalence), mean bias was negligible, mean confidence interval width was acceptable and the proportion of valid estimates was >95% for all pool sizes up to 20 and for both perfect tests and for an assumed test sensitivity of 0.9 instead of 1.0.

Table 4: Summary results of simulated pooled testing for three case studies, including number of pools tested, mean bias, mean confidence interval width and proportion of valid estimates for pool sizes of 2, 3, 4, 5, 10 and 20, and for both a perfect test and a test with known sensitivity of 0.9 and specificity of 1.

		Case Study 1: prevalence = 0.59, precision = 0.04				Case Study 2: prevalence = 0.14, precision = 0.055				Case Study 3: prevalence = 0.016, precision = 0.014			
Strategy	Pool size	Number of pools	Mean bias	Mean CI width	Proportion Valid	Number of pools	Mean bias	Mean CI width	Proportion Valid	Number of pools	Mean bias	Mean CI width	Proportion Valid
Sensitivity = Specificity = 1													
1	2	499	-0.001	0.082	0.960	83	0.000	0.115	0.960	156	0.000	0.031	0.980
2	3	606	0.002	0.083	0.953	60	0.000	0.116	0.959	105	0.000	0.031	0.978
3	4	867	0.001	0.084	0.961	59	0.002	0.117	0.960	79	0.000	0.031	0.991
4	5	N/A	N/A	N/A	N/A	42	0.002	0.119	0.965	64	0.000	0.032	0.967
5	10	N/A	N/A	N/A	N/A	33	0.006	0.129	0.958	33	0.000	0.032	0.977
6	20	N/A	N/A	N/A	N/A	46	0.079	0.208	0.974	18	0.001	0.033	0.968
Sensitivity = 0.9, Specificity = 1													
1	2	341	0.000	0.129	0.966	82	0.001	0.126	0.975	171	0.000	0.032	0.985
2	3	147	0.009	0.356	0.973	47	0.001	0.145	0.967	112	0.000	0.032	0.980
3	4	78	-0.018	0.583	0.773	30	0.003	0.171	0.982	83	0.000	0.033	0.968
4	5	49	-0.068	0.681	0.613	20	0.004	0.218	0.971	65	0.000	0.033	0.981
5	10	14	-0.362	0.893	0.419	5	-0.007	0.830	0.830	29	0.001	0.037	0.977
6	20	4	-0.511	0.981	0.348	N/A	N/A	N/A	N/A	12	0.001	0.049	0.984

The most frequent result (number of positive pools) and corresponding estimated prevalence and confidence limits for selected pool sizes for each case study are summarised in Table 5. For all case studies and sensitivity levels, the estimated prevalence was close to the true prevalence and the true value was well within the estimated 95% confidence limits. Exact confidence interval widths were close to the target values for an assumed test sensitivity of 1.0 for all cases and for an assumed test sensitivity of 0.9 for the low prevalence case (Case Study 3). For an assumed test sensitivity of 0.9 and high or moderate prevalence (Case Studies 1 and 2), confidence interval widths were larger than predicted from the sample size calculation, because the asymptotic methods used for sample size calculation underestimate confidence interval width compared to exact methods when an imperfect test is assumed.

Table 5: Most frequent results for selected pool sizes and corresponding prevalence and confidence limit estimates for each of three case studies and for assumed test sensitivities of 0.9 and 1.0 and specificity of 1.0

Case Study	Assumed Sensitivity/ Specificity	Pool size	Number of pools tested	Number of positive pools	Estimated prevalence	Exact 95% Confidence limits	Confidence interval width
1	1.0/1.0	2	499	415	0.590	0.548 - 0.631	0.083
1	0.9/1.0	2	341	256	0.593	0.530 - 0.660	0.13
2	1.0/1.0	5	42	22	0.138	0.083 - 0.193	0.11
2	0.9/1.0	5	20	10	0.150	0.069 - 0.282	0.213
3	1.0/1.0	10	33	5	0.016	0.005 - 0.038	0.033
3	0.9/1.0	10	29	3	0.012	0.003 - 0.036	0.033

11.4. Discussion

These case studies clearly demonstrate the savings in the numbers of tests required that can be achieved by pooled testing. The first analysis showed that for the original samples, pooling of the original samples prior to testing would have provided an unbiased estimate of the observed prevalence, with comparable confidence limits to the original unpooled testing. The maximum appropriate pool size that could be used depended on the assumed prevalence, ranging from two samples per pool for the high-prevalence (59%) situation up to 29 (or more) samples per pool for the low-prevalence scenario (1.6%). These results assume no effect of dilution, and that the test had 100% sensitivity and specificity. However, adjusting for possible loss of sensitivity due to pooling (assumed to be reduced to 90%) resulted in similar results.

Similarly, except for the high-prevalence situation, the number of tests required to provide equivalent confidence and precision was lower for pooled testing than for individual testing, although the total number of samples to be collected was higher, assuming a perfect test. For the high-prevalence situation, the very high prevalence observed in this population meant that for all pool sizes > 2, nearly all pools tested positively, so that a very large number of pools would have to be tested to be confident of obtaining sufficient negative pools to allow prevalence to be estimated.

If the test was assumed to have imperfect sensitivity (Se = 90%) for pooled testing the resulting number of pools was generally smaller for each pool size than for the same pool size with a perfect test. This was due to the effect of test sensitivity being applied at the pool level, so that even in the high-prevalence situation, fewer pools tested positively, reducing bias and improving confidence. However, these sample sizes were based on large sample theory and

asymptotic confidence limits. Subsequent simulation using these sample sizes showed that the exact confidence limits were wider than expected because of this difference.

These case studies also highlighted an important potential pitfall of pooled testing for prevalence estimation – estimates can be substantially biased if prevalence is high and pool size is too large. This was particularly apparent in the high-prevalence situation, where pool sizes >2 resulted in increasing bias in the prevalence estimate and increasing loss of precision in the confidence limits about the estimate. It was also apparent to a lesser extent at moderate prevalence, where a pool size of 20 resulted in a substantial increase in bias and confidence-interval width. In contrast, there was no evidence of bias or loss of precision at any pool size considered in the low-prevalence case study.

11.5. Conclusion

Based on these case studies, pooled testing is likely to be a useful tool for estimating prevalence of new and emerging diseases such as Hendra or Nipah virus, particularly where the true prevalence is expected to be low-moderate. Even in high prevalence situations, pooled testing may be useful if small pool sizes are used.

However, there are a number of important issues that must be considered when designing any pooled testing survey. Firstly, the selection of an appropriate pool size and number of pools is essential, and should be based on consideration of the expected true prevalence of disease and the desired precision and confidence in the estimate.

Secondly, it is essential that samples are selected and pooled in as random a manner as possible. Clustering of infection in some sub-populations is likely, and if sampling and pooling is not undertaken at random it is likely that this clustering will also occur within the final pools, resulting in seriously biased estimates.

Finally, simulation provides an important tool for assessing alternative pooling strategies to identify the appropriate strategy that meets the survey design parameters with an acceptably low level of bias and at an acceptable cost.

11.6. References

Sahani M, Parashar UD, Ali R, Das P, Lye MS et al., 2001. Nipah virus infection among abattoir workers in Malaysia 1998-1999. *Int. J. Epidemiol.* 30: 1017-1020.

12. Appendix 3: Case study – Johne's disease

Australian Biosecurity CRC for Emerging Infectious Disease

Project Number	**3.004 R**
Project Title	**Estimation of prevalence from pooled samples**

Report on

Application of Software to Ovine Johne's Disease Data

Prepared by
Jenny-Ann Toribio

12.1. Background

Research under Australian conditions on the efficacy of a killed whole cell vaccine (Gudair ®, CZ Veterinaria, Spain) commenced in 1999. To be registered for disease control use in Australia, the Gudair® vaccine had to be shown to significantly reduce the prevalence of faecal shedding of *Mycobacterium avium* subsp. *paratuberculosis* (MAP) and mortalities in sheep exposed to high levels of MAP over their lifetime. However, a longitudinal on-farm trial using only individual faecal culture was untenable due to high cost.

Consequently, a 4-year vaccine trial has used pooled faecal culture, with reduction in pool size from 40 to 10 as disease progressed, to achieve prevalence estimates with adequate confidence intervals and to reduce project costs for laboratory work. The main objectives of this trial were to determine whether vaccination reduces the incidence of mortality and the faecal shedding of MAP in terms of the prevalence of shedders and excretion levels.

In the trial, conducted on three properties in southern New South Wales with OJD losses (5 to 15% per annum), 200 Merino lambs were vaccinated with Gudair®, and 200 with a placebo saline vaccine in each flock at 2 to 3 months of age. These animals were exposed to high levels of MAP throughout the trial by being run together with OJD clinical adult sheep. Samples and data were collected twice yearly to assess the effect of vaccination on faecal shedding of MAP, mortality rate, growth, wool production, vaccine injection site lesions, and cellular and humoral immunity.

At each observation, faeces were collected from individual sheep, pooled and cultured for MAP. At trial commencement the pool size was 40 sheep (5 vaccinate pools and 5 control pools were cultured per flock). This pool size reduced to 20 (10 vaccinate pools and 10 control pools) and then to 10 (20 vaccinate pools and 20 control pools) as the number of MAP shedders increased with disease progression. In addition at samplings conducted since December 2000 individual faecal samples were collected and stored at -80^{o} C. Culture of these individual samples was conducted when the respective pool tested positive for MAP or when the respective pool tested negative and was among four negative pools per observation randomly selected for individual culture.

Pooled culture results were used to estimate individual animal prevalence among vaccinate and control sheep. To date prevalence of shedding in each group based on pooled faecal culture (PFC) has been estimated using two separate methods (methods 2 and 6) as defined by Cowling et al. (1999). The individual culture results were used to verify the prevalence estimates obtained from PFC.

Selection of the preferred method to estimate individual-level prevalence from pooled cultures depends on factors such as pool size, number of pools, knowledge of test sensitivity and specificity, and true disease prevalence. So far in this study two methods have provided point and confidence interval estimates of prevalence. However, neither is entirely appropriate due to reliance on large-sample theory, requirement for fixed pool size and lack of correction for test sensitivity (Method 2 assumes perfect sensitivity and specificity; Method 6 adjusts for imperfect sensitivity but this is limited by current knowledge about PFC sensitivity at various pool sizes). To obtain more accurate prevalence estimates, it is evident that analysis of the PFC data should be extended to other methods including a Bayesian approach.

Alternate methods for estimation of individual animal prevalence from pooled results are documented in the literature (Sacks et al., 1989; Cowling et al., 1999; Williams and Moffitt, 2001). Recently Evan Sergeant developed computer software to implement several of these methods in a user-friendly application known as the *Pooled Prevalence Calculator (PPC)*. Use of this software to analyse the trial data will permit estimation of shedding prevalence by up to seven different methods. Comparison of the results for each method will identify the method that provides the most accurate estimates of prevalence and related confidence intervals. This document reports application of the PPC to the pooled data from the Gudair vaccine trial.

12.2. Objectives

The objectives of the work documented in this report were:

1. To test the functionality of specific components in the Pooled Prevalence Calculator
2. To estimate individual animal prevalence of shedders in vaccinate and control groups by several alternate methods
3. To identify the method that provides the most accurate estimates of prevalence and related confidence intervals.

12.3. Methods

From raw data on the faecal culture results for all observations of trial sheep to December 2002[1], a spreadsheet was constructed in Excel to record separately for vaccinate and control groups at each observation:

- Date of observation
- Number of months post-vaccination
- Total number of sheep sampled

For pooled faecal cultures

- Pool size
- Number of pools tested for each pool size
- Number of positive pools for each pool size

For individual faecal cultures

- Number of sheep with IFC
- Number of sheep with positive IFC.

Features of this dataset with implication for analyses include:

- Presence of estimates rather than definite total numbers of sheep sampled for observations prior to June 2002
- Results for pools of various sizes (rather than one pool size) present for observations from June 2002 – for all analyses assuming a fixed pool size, all pools were assumed to be equal to the most frequent pool size.
- Number of sheep with positive IFC results for each observation is based on individual culture of sheep represented in all positive pools and four randomly selected negative pools.

Due to variation in the dataset (particularly in relation to pool size) and lack of published figures for the sensitivity of PFC at various pool sizes, none of the methods for estimating

[1] Provided by Dr Leslie Reddacliffe

individual prevalence from pooled samples present in the PPC software is completely appropriate for the trial data. Thus assumptions were made to permit the calculation of individual prevalence estimates (and related confidence intervals) by six of the seven methods in the PPC. Further the estimation of individual prevalence was restricted to one observation period (with sheep 30 months post-vaccination) when the target pool size on all three properties was 10. For this pool size, unlike pool sizes of 40 and 20, published sensitivity and specificity figures were available for the BACTEC PFC (Whittington et al., 2000)[2].

Table 1 presents the assumptions and methods used (in relation to each assumption) to produce individual prevalence estimates by six methods in the PPC. When fixed pool size was assumed, prevalence estimates were produced using methods 3, 4, 5 and Gibbs sampler (which differ in assumptions related to test sensitivity and specificity). When PFC was assumed to have 100% sensitivity and specificity, prevalence estimates were produced using method 1 and the variable pool size method.

The Gibbs sampler method requires inputs for prior estimates of the true prevalence and test sensitivity and specificity, based on expert knowledge or previous data (Sergeant, 2004). Two prior estimates of true prevalence based on culture results of the prior observation conducted at 23-24 months post-vaccination were used. The first (identified as B1 in figures) was based on the number of IFC positive sheep and the total number of sheep at this observation and the second (identified as B2 in figures), the number of PFC positive pools and the total number of pools at this observation.

For the first prior estimate of true prevalence, the Beta probability distribution parameters alpha (α) and beta (β) were calculated as follows:

$$\alpha = x + 1$$
$$\beta = n - x + 1$$

x - the number of IFC positive sheep at the prior observation
n - total number of sheep sampled at the prior observation

For the second prior estimate of true prevalence and for the prior estimates of test sensitivity and specificity, the Beta probability distribution parameters alpha (α) and beta (β) were calculated using the Beta distribution utility in the PPC. To calculate parameters for prior probability distributions this PPC function requires inputs for mode and 5/95 percentile based on expert knowledge or previous data. These input figures for the second prior estimate of true prevalence were based on the individual estimates produced by method 5 on entry of pool data results from the prior observation. For test sensitivity and specificity, the mode and 5/95 percentile figures were based on results published for pools of 10 by Whittington et al. (2000).

In addition, point prevalence of shedders at 30 months post-vaccination was calculated using the IFC results (proportion x/n) for comparison with point prevalence estimates produced by the 7 PFC methods.

When either Bayesian approach (B1 or B2) produced point estimates notably higher than all other methods (including prevalence based on IFC results), further calculations were

[2] On the basis of sensitivity estimates for multibacillary and paucibacillary cases in pools of 10 of 100% and 60% (Whittington et al., 2000) and an assumption of equal proportions of multibacillary and paucibacillary cases in the trial sheep, the sensitivity was set at 80%. PFC specificity for pools of 10 was set at 100%.

performed with modified prior distributions for prevalence to identify the potential source of the variation.

Table 1
Method used to calculate individual prevalence in the PPC

Assumption 1	Assumption 2	Method
Fixed pool size	Known test sensitivity and specificity[a]	Method 3 Method 4
	Unknown test sensitivity and specificity[b]	Method 5
	Estimates of prior Beta distributions for true prevalence and test sensitivity and specificity[c]	Pooled prevalence using a Gibbs sampler
Perfect test	Fixed pool size	Method 1
	Variable pool size	Pooled prevalence for variable pool size and perfect tests

a Assuming fixed pool size of 10 and equal proportions of multibacillary and paucibacillary cases, sensitivity was set at 80% and specificity at 100%.
b On the basis of the same assumptions, sensitivity was set at 80% and specificity at 100%. In addition the sample size for estimating sensitivity (ν_1) was set at 50 and the sample size for estimating specificity (ν_2) at 5000.
c Parameters used for the two prior estimates of true prevalence are shown in Table 3. Parameters used for sensitivity, based on mode of 0.8 and 5/95 percentile of 0.6, were α of 14.84 and β of 4.46, and for specificity, based on mode of 0.999 and 5/95 percentile of 0.98, were α of 163.44 and β of 1.16. The α and β values were calculated using the Beta distribution utility in the PPC.

12.4. Results

At the observation completed at 30 months post-vaccination, faecal samples were collected from 173 vaccinates and 162 controls on Property 1, 169 vaccinates and 144 controls on Property 2, and 183 vaccinates and 167 controls on Property 3. These faecal samples were cultured in pools of various size as presented in Table 2. Subsequently faeces from sheep represented in positive pools were individually cultured and IFC results were positive for 1 vaccinate and 5 controls on Property 1, 4 vaccinates and 23 controls on Property 2, and 5 vaccinates and 29 controls on Property 3. Table 3 presents estimates of the point prevalence of shedders based on these IFC results for vaccinate and controls groups on each property.

For the prior estimates for PFC test sensitivity and specificity (for pools of 10), the alpha and beta parameters calculated were 14.84 and 4.46, respectively for sensitivity (with mode = 0.8 and 5/95 percentile = 0.6 inputs), and 163.44 and 1.16, respectively for specificity (with mode = 0.999 and 5/95 percentile = 0.98 inputs).

Table 4 presents the alpha and beta parameters calculated for the two prior estimates of true prevalence based on culture results for the previous observation at 23-24 months post-vaccination.

Estimates of the prevalence of shedders in each trial group at 30 months post-vaccination based on PFC and IFC are shown in Figures 1 to 6.

Bayesian approaches produced substantially higher point prevalence estimates in three cases (Figures 3, 4 and 6). Recalculations in each case using modified priors was performed (Table 5) and resulted in a reduction in the point estimate for the controls on Property 2 (from 44.1% to 30.3%) and Property 3 (from 26.5% to 20.3%), but no change for the vaccinates on Property 2.

Table 2
Number of pools tested by PFC and number positive by pool size for faecal samples collected from vaccinate and control groups on 3 properties 30 months post-vaccination

Property / Trial group	Pool size													
	10		9		8		7		5		4		3	
	No. pools	No. positive	No. pools	No. positive	No. pools	No. positive	No. pools	No. positive	No. pools	No. positive	No. pools	No. positive	No. pools	No. positive
1														
Vaccinates	15	1	2	0					1	0				
Controls	10	5	9	6	1	0								
2														
Vaccinates	16	4	1	0										
Controls	14	11									1	1		
3														
Vaccinates	18	3											1	1
Controls	16	13					1	1						

Table 3

Point prevalence based on IFC results for vaccinate and controls groups on 3 properties 30 months post-vaccination

Property / Trial group	Total samples	Number IFC positive	IFC prevalence (%)
1			
Vaccinates	173	1	0.6
Controls	162	5	3.1
2			
Vaccinates	169	4	2.4
Controls	144	23	16.0
3			
Vaccinates	183	5	2.7
Controls	167	29	17.4

Table 4
Parameters for two prior estimates of true prevalence in vaccinate and control groups on 3 properties based on faecal culture results 23-24 months post-vaccination

Property / Trial group	Based on IFC						Based on PFC			
	x	n	α	β	Mode	5/95 percentile	Mode	5/95 percentile	α	β
1										
Vaccinates[ab]	0	180	1	181	0	0.016				
Controls[a]	8	180	9	173	0.044	0.078	0.042	0.08	7.16	141.51
2										
Vaccinates	6	187	7	182	0.032	0.062	0.049	0.091	7.6	129.09
Controls[c]	80	182	81	103	0.44	0.501				
3										
Vaccinates	2	184	3	183	0.011	0.034	0.014	0.034	4.42	241.87
Controls	23	173	24	151	0.133	0.182	0.193	0.397	4.21	14.45

a The figures of 180 vaccinates and 180 controls sampled on Property 1 are estimates.

b Pool size of 20 used for vaccinates at previous observation. Estimates of sensitivity and specificity not available for this pool size.

c Method 5 unable to provide mode and percentile figures due to 18/19 pools being positive for controls at previous observation.

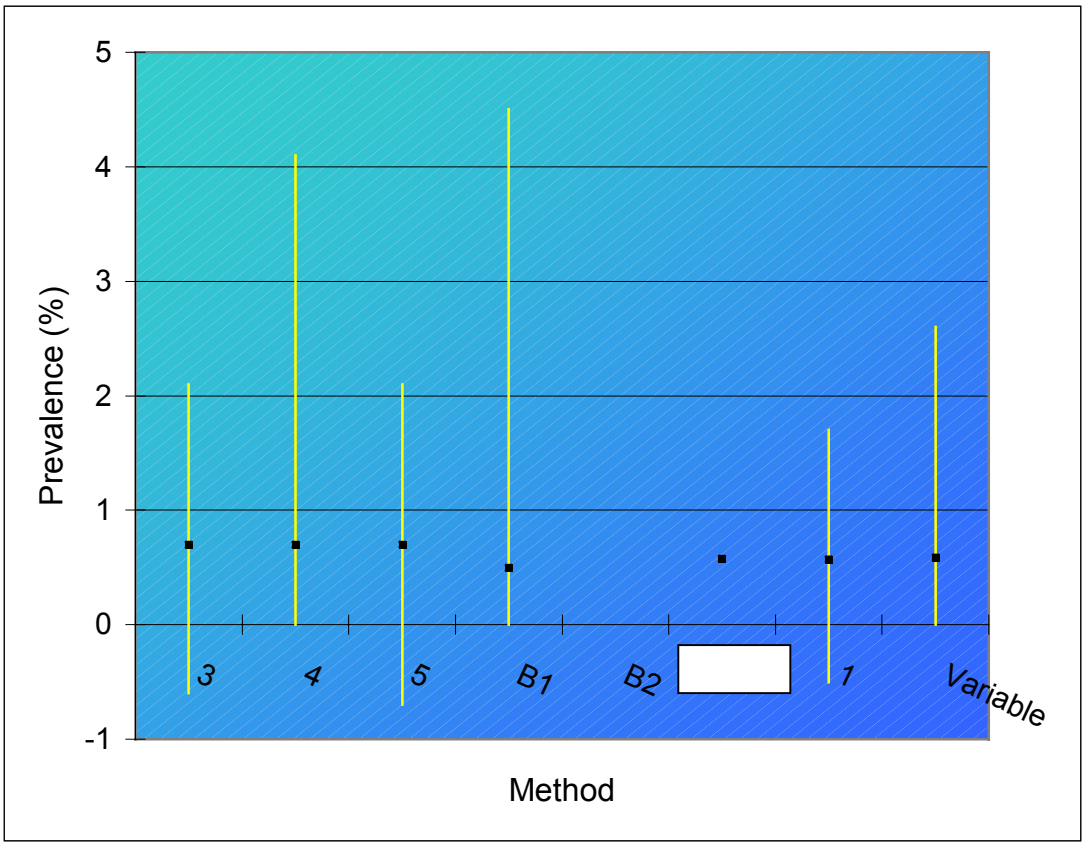


Figure 1
Estimates of the prevalence of shedders among vaccinates on Property 1 at 30 months post-vaccination calculated by 6 methods using the PPC and IFC. Due to lack of estimates for sensitivity and specificity for pools of 20, no priors for prevalence were available for method B2.

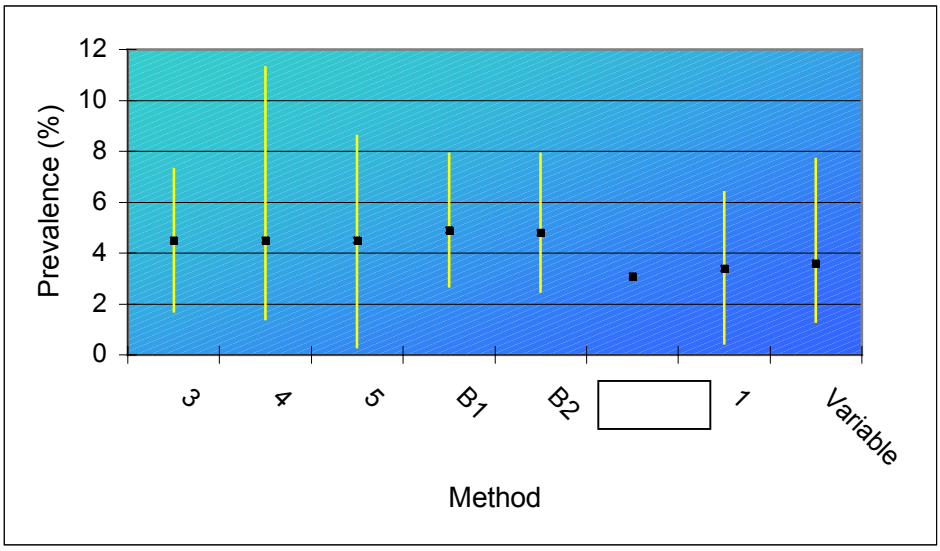


Figure 2
Estimates of the prevalence of shedders among controls on Property 1 at 30 months post-vaccination calculated by 7 methods using the PPC and IFC.

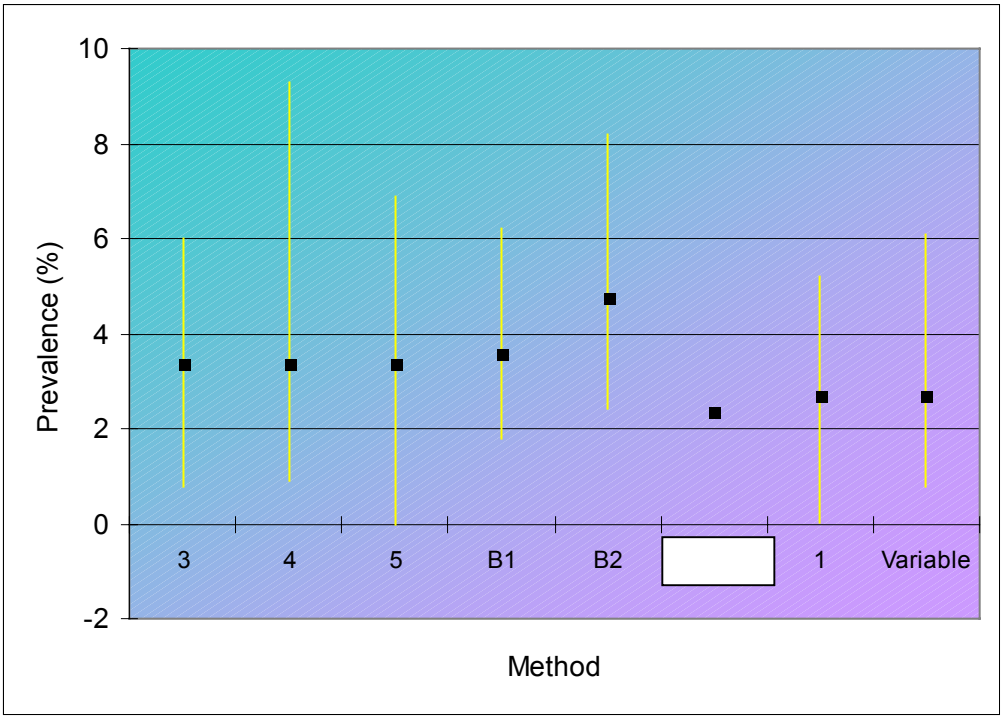


Figure 3
Estimates of the prevalence of shedders among vaccinates on Property 2 at 30 months post-vaccination calculated by 7 methods using the PPC and IFC.

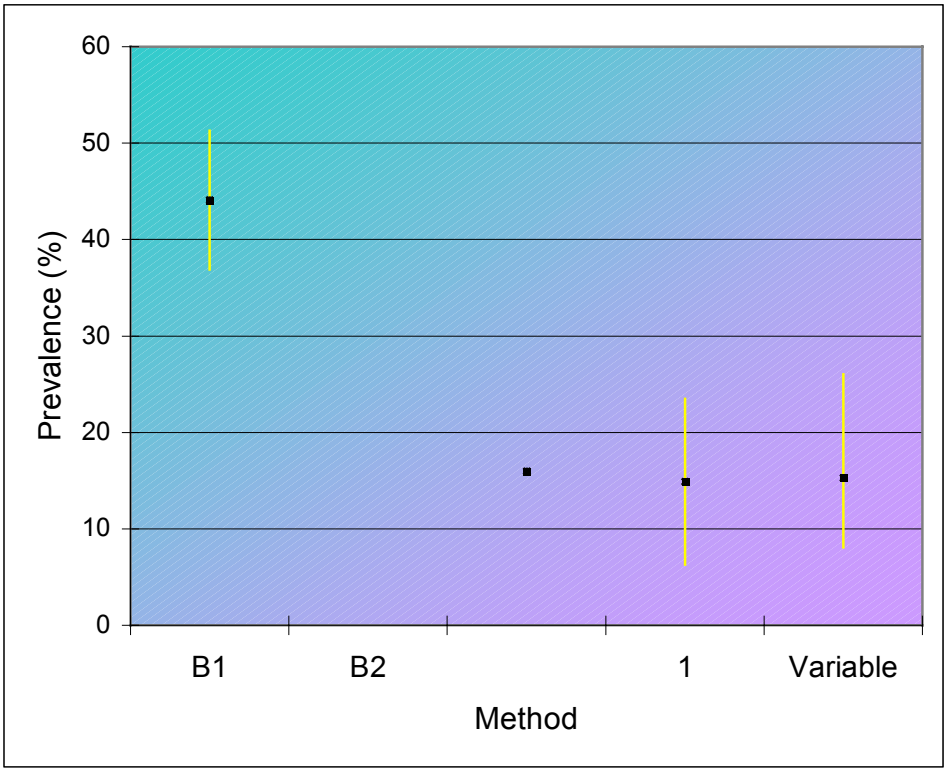


Figure 4
Estimates of the prevalence of shedders among controls on Property 2 at 30 months post-vaccination calculated by 3 methods using the PPC and IFC.
Methods 3, 4 and 5 could not compute prevalence estimates due to high proportion of positive pools (12/14) at 30 months post-vaccination. Likewise Method 5 could not compute mode and percentile figures at 23-24 months post-vaccination (18 positive /19 pools) so no priors for prevalence were available for method B2.

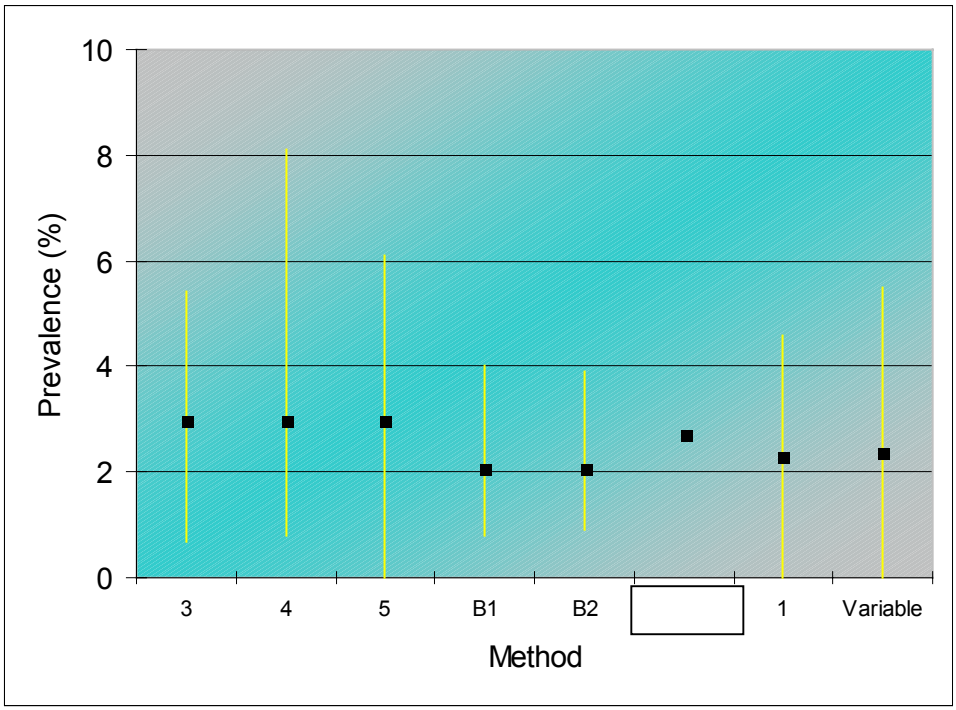


Figure 5
Estimates of the prevalence of shedders among vaccinates on Property 3 at 30 months post-vaccination calculated by 7 methods using the PPC and IFC.

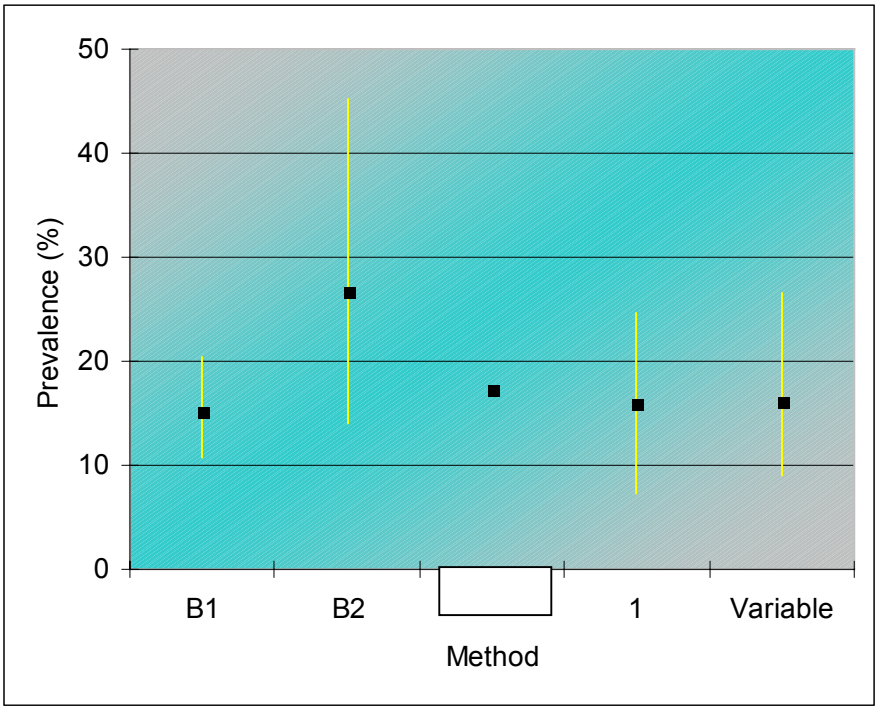


Figure 6
Estimates of the prevalence of shedders among controls on Property 3 at 30 months post-vaccination calculated by 4 methods using the PPC and IFC.
Methods 3, 4 and 5 could not compute prevalence estimates due to high proportion of positive pools (14/16) at 30 months post-vaccination.

Table 5
Estimates of point prevalence for 3 trial group at 30 months post-vaccination using modified priors for prevalence

Property / Trial group	Mode	5/95 percentile	α	β	Prevalence		
					LCL	Mean	UCL
2 Vaccinates	0.049	0.091	7.60	129.09	2.4	4.8	8.3
	0.049	0.12	3.99	59.03	2.0	4.8	9.3
	0.049	0.15	2.81	36.13	1.9	4.9	10.1
	0.049	0.20	2.03	20.99	1.6	4.8	11.6
2 Controls	0.44	0.501	81	103	36.9	44.1	51.3
	0.40	0.47	56.45	84.17	32.0	40.2	48.3
	0.35	0.42	48.66	89.51	27.6	35.2	43.6
	0.30	0.37	40.29	92.68	23.1	30.3	38.3
3 Controls	0.193	0.397	4.21	14.45	13.7	26.5	46.1
	0.193	0.35	6.07	22.24	13.2	24.1	39.5
	0.193	0.30	10.93	42.6	13.8	22.3	33.1
	0.193	0.25	31.59	129.15	14.9	20.3	26.6

12.5. Discussion

The components of the PPC used to complete this report proved to be functional and to provide results in tabular and graphical formats helpful for the user. Two useful additions to the website would be:

- a complete user guide in pdf format for download by users
- 5th and 95 th percentile figures in the output from the Beta probability distribution utility for specified alpha and beta parameters.

Each method for estimation of individual prevalence from pooled samples included in the PPC functioned in accordance with its model specifications and demonstrated its inherent limitations. For example, methods 1, 3 and 5 apply large-sample theory and this resulted in a negative lower confidence limit for the low OJD prevalence vaccinate group on Property 1. Conversely, prevalence estimates could not be computed for the high OJD prevalence control groups on Properties 2 and 3 by methods 3, 4 and 5 due to the proportion of positive pools exceeding the nominated test sensitivity. In addition the advantage of the Bayesian method, computation of narrower confidence intervals, was demonstrated in all but two instances. It is likely that the exceptions result from inaccurate prior estimates for prevalence due to reliance on estimates at 23-24 months post-vaccination based on zero positive IFC from an estimated total of 180 sheep (for vaccinates on Property 1) and on method 5 calculations when the proportion of positive pools (12/17 pools or 70.5%) approached the nominated test sensitivity of 80% (for controls on Property 3). However, it is also important to realise that the Bayesian estimates are affected by the prior estimates used, and that if these estimates or the width of the distribution used are inappropriate the resulting estimate could be biased and the probability interval unrealistically narrow.

Comparison of point prevalence estimates calculated from IFC results and the point estimates produced by each of the PPC methods showed the following:

- With two exceptions, the IFC estimate was within the 95% confidence/credibility of the pooled estimate for all properties and methods. The exceptions were the vaccinates on Property 2 using method B2 and the controls on Property 2 using method B1 (no priors for prevalence were available for method B2 for the controls). For the vaccinates, the median and lower 95% confidence limit were substantially higher than equivalent measures for all other methods. For the controls, the median and upper and lower confidence limits were dramatically higher than equivalent measures for all other methods. These outlying results suggest that the Bayesian estimates in each instance are possibly biased due to an inappropriate prior distribution.
- In general, the IFC estimate was close to the point estimates from methods assuming a perfect test, and well within 95% confidence limits for these methods
- In four cases the IFC point estimate was closest to point estimates calculated by method 1 and by the method for variable pool size - both methods that assume a perfect test.
- For Methods 3, 4 and 5 (frequentist methods for an imperfect test) , where they were possible, estimates were generally higher than the IFC estimate, although the IFC estimate was within the 95% confidence limit on all occasions.
- Bayesian methods tended to produce narrower confidence intervals than the comparable frequentist methods and varied in comparison to the IFC and other pooled estimates, mainly because of the effect of the prior prevalence distributions used.

Notably the calculation of IFC prevalence used assumed 100% sensitivity and 100% specificity in contrast to the reported variation in sensitivity for IFC of 100% for multibacillary sheep and 70% for paucibacillary sheep (Whittington et al., 2000). IFC prevalence would therefore be expected to be similar to pooled estimates assuming perfect tests and to underestimate prevalence compared to pooled estimates using methods which adjust for imperfect sensitivity.

The second Bayesian approach (B2) produced point estimates higher than other methods (and substantially higher than IFC) in two cases – vaccinates on Property 2 and controls on Property 3. In addition the first Bayesian approach (B1) produced a high outlying point estimate for controls on Property 2. These results are most likely related to the choice of prior distributions for prevalence in these analyses (both mode and width of distribution), and demonstrates the importance of selecting an appropriate prior distribution for the Bayesian analysis. For the two cases involving the B2 method, the prevalence priors based on PFC results at 23-24 months post-vaccination differed from those based on IFC. Recalculation with alpha and beta parameters based on adjusted mode and/or 5/95 percentile figures produced reduced point estimate and/or confidence limits. This demonstrates the influence of the prevalence priors on the output from the Bayesian method particularly when dealing with small numbers of pool samples. For example, the outlying point estimate for the Property 2 controls tends to reflect the high prior based on IFC for 182 sheep 23 months post-vaccination more than the PFC results for 144 sheep at 30 months post-vaccination.

The PPC presents current methods for individual prevalence estimation from pooled samples in a user-friendly format. However none of these methods is adequate to address the complexity of the OJD trial data. Although none of the methods is ideal, it is important to note that for all but one of the analyses carried out, the 95% confidence/credibility interval included the value for prevalence estimated from IFC. In addition, the IFC estimate is itself biased, because of the assumption of perfect sensitivity and specificity of IFC, and the fact that not all negative pools were tested by IFC. Therefore, the IFC estimate is also likely to underestimate true prevalence, and adjustment for these effects should bring the IFC estimate even closer to the estimates generated using methods that account for imperfect test performance.

Further research is needed to develop a method that will account for variation in pool size and in test sensitivity and specificity at different pool sizes and prevalence levels. More investigation is also required to determine the best method for identifying an appropriate prior distribution for prevalence in situations such as this and to determine the performance characteristics of PFC at various pool sizes with different proportions of multibacillary and paucibacillary sheep.

12.6. Response to issues raised

This report raises a number of issues worthy of response or further comment:

1. Suggested changes to the Pooled Prevalence Calculator

Suggestions for a downloadable pdf version of the User Guide and for addition of 5 and 95 percentile figures to the output from the Beta probability distribution utility have both been implemented.

2. Results for Bayesian methods

The results reported for the Bayesian analyses highlight the importance of selection of an appropriate prior distribution. Based on these results it appears that the prior distributions used placed too much reliance on previous testing results, in a situation where prevalence in the flocks apparently changed quite dramatically between tests in some instances. In this case much weaker priors or perhaps a uniform prior would have been preferable, and these alternatives are planned for further investigation.

12.7. References

COWLING, D. W., GARDNER, I. A. and JOHNSON, W. O. (1999). Comparison of methods for estimation of individual-level prevalence based on pooled samples. *Preventive Veterinary Medicine* **39,** 211-225.

SACKS, J. M., BOLIN, S. and CROWDER, S. V. (1989). Prevalence estimation from pooled samples. *American Journal of Veterinary Research* **50,** 205-206.

SERGEANT, E. (2004). Pooled Prevalence Calculator, AusVet Animal Health Services. http://www.ausvet.com.au/pprev/

WHITTINGTON, R. J., FELL, S., WALKER, D., MCALLISTER, S., MARSH, I., SERGEANT, E., TARAGEL, C. A., MARSHALL, D. J. and LINKS, I. J. (2000). Use of pooled faecal culture for sensitivity and economic detection of Mycobacterium avium subsp. Paratuberculosis infection in flocks of sheep. *Journal of Clinical Microbiology* **38,** 2550-2556.

WILLIAMS, C. J. and MOFFITT, C. M. (2001). A critique of methods of sampling and reporting pathogens in populations of fish. *Journal of Aquatic Animal Health* **13,** 300-309.