Medical and Veterinary Entomology

Medical and Veterinary Entomology (2012) **26**, 361–371 doi: 10.1111/j.1365-2915.2012.01015.x

REVIEW ARTICLE

Estimating the prevalence of infections in vector populations using pools of samples

N. SPEYBROECK[1], C. J. WILLIAMS[2], K. B. LAFIA[3], B. DEVLEESSCHAUWER[1,4] and D. BERKVENS[5]

[1]Institut de Recherche Santé et Société (IRSS), Université catholique de Louvain, Brussels, Belgium, [2]Department of Statistics, University of Idaho, Moscow, ID, U.S.A., [3]Faculty of Agricultural Sciences, University of Abomey Calavi, Cotonou, Benin, [4]Department of Virology, Parasitology and Immunology, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium and [5]Department of Animal Health, Institute of Tropical Medicine, Antwerp, Belgium

Abstract. Several statistical methods have been proposed for estimating the infection prevalence based on pooled samples, but these methods generally presume the application of perfect diagnostic tests, which in practice do not exist. To optimize prevalence estimation based on pooled samples, currently available and new statistical models were described and compared. Three groups were tested: (a) Frequentist models, (b) Monte Carlo Markov-Chain (MCMC) Bayesian models, and (c) Exact Bayesian Computation (EBC) models. Simulated data allowed the comparison of the models, including testing the performance under complex situations such as imperfect tests with a sensitivity varying according to the pool weight. In addition, all models were applied to data derived from the literature, to demonstrate the influence of the model on real-prevalence estimates. All models were implemented in the freely available R and OpenBUGS software and are presented in Appendix S1.

Bayesian models can flexibly take into account the imperfect sensitivity and specificity of the diagnostic test (as well as the influence of pool-related or external variables) and are therefore the method of choice for calculating population prevalence based on pooled samples. However, when using such complex models, very precise information on test characteristics is needed, which may in general not be available.

Key words. Arthropod vector, model, pool of samples, prevalence.

Introduction

Contemporary phenomena such as globalization, global warming, changing land use and increased water storage, population growth and urbanization are causing arthropod disease vectors, such as mosquitoes, ticks and sandflies, to be dispersed from their initial biotopes over the whole world (Sutherst *et al.*, 1998). The diseases transmitted by these vectors are becoming a disquieting public and animal health problem (Gratz, 1999), which needs to be quantified and monitored.

The prevalence of any arthropod-borne pathogen in a population of vectors reflects the level of transmission of the pathogen and is a key factor in the epidemiology and risk analysis of the corresponding disease. Unfortunately, it is not easy to determine the prevalence in vector populations. As the prevalence is often lower than 10% (Katholi *et al.*, 1995; Abel *et al.*, 1999), high numbers of vectors need to be collected, and each individual vector in this sample would then need to be subjected to one or more diagnostic tests to determine its infection status (Munoz-Zanzi *et al.*, 2006). The resulting high diagnostic costs necessitated the development of alternative solutions. The screening of pools or clusters of vectors was first used over 60 years ago, as reported by Dorfman (1943). Since then, it has become current practice in mass screening exercises for medical and veterinary pathogens (Farrington, 1992; Mendoza-Blanco *et al.*, 1996; Cowling *et al.*, 1999; Martin-Sanchez *et al.*, 2006).

In certain instances, if a pool is detected as positive, it would be possible to go back to the original constituent samples and directly count the number of positives and estimate the

Correspondence: Niko Speybroeck, Institut de Recherche Santé et Société (IRSS), Université catholique de Louvain, Clos Chapelle-aux-Champs, 30 boîte 3005, 1200 Brussels, Belgium. Tel.: + 32 2 764 3375; Fax: + 32 2 764 3470; E-mail: Niko.Speybroeck@UCLouvain.be

individual-based prevalence. However, because of financial or practical constraints, this may not always be possible. Statistical methods that allow estimating the individual-based prevalence based on the pool prevalence are therefore needed. The most straightforward approach would be to divide the number of positive pools by the total number of samples tested [thereby obtaining an estimate of the Minimum Infection Rate (MIR)]. However, as the MIR can underestimate the prevalence, several other statistical methods have been proposed to estimate the individual-based prevalence based on the pool prevalence (Farrington, 1992; Abel *et al.*, 1999; Cowling *et al.*, 1999).

A challenge when estimating the prevalence of an infection in arthropod vector populations, as in any population, is the lack of reliable tests (Watt *et al.*, 1997). Perfect diagnostic tests do not exist, especially when dealing with field collections. Several diagnostic tests exist for studying infections in vectors ranging from histological tests to serological tests and polymerase chain reaction (PCR) tests (Katholi *et al.*, 1995). The latter allows for the detection of the disease-causing agent's DNA in the vectors, and is considered to have high sensitivity and specificity. Polymerase chain reaction tests may be complemented using restriction fragment length polymorphism or gene sequencing approaches to identify the detected species, thereby increasing the test specificity even further (Yameogo *et al.*, 1999; Oshaghi *et al.*, 2010). However, the sensitivity and specificity of any test, including the PCR test, can be influenced by several endogenous and exogenous factors, and can thus not be considered as a constant provided by the diagnostic test manufacturer (Berkvens *et al.*, 2006; Speybroeck *et al.*, 2011). For example, the test sensitivity can change according to the presence of inhibitors in the samples to be tested. During bloodmeals, arthropod vectors engorge themselves, which may bring about inhibiting effects on the TAQ polymerase, an enzyme intervening in the PCR process (Schwartz *et al.*, 1997; Al-Soud *et al.*, 2005). It is therefore expedient to have statistical techniques that can account for these imperfect and variable test characteristics.

The present study describes and compares existing and new models for estimating the prevalence in haematophagous arthropod populations, and explores how models can account for the effect of modulating factors on the test characteristics, using the inhibition example.

Materials and methods

The present study will:

- describe methods for estimating infection prevalence in vector populations based on pooled samples,
- synthesize existing computer code for these models and, where necessary, present new code,
- adjust the various models, where possible, to account for imperfect tests with and without dependence on modulating factors, exploring e.g. the feasibility of the diagnostic test sensitivity to be a function of endogenous or exogenous factors, such as inhibition,
- test the various models using data generated according to predefined prevalences and test characteristics,
- apply the models to data from the literature and
- compare the performance of the various models.

Statistical formulations

The following notations will be used:

- p: individual-based prevalence
- k: size of the pools (k_j if the pool sizes vary)
- P_T: true probability of having at least one vector positive in the pool; $P_T = 1 - (1 - p)^k$
- SE: sensitivity of the test for diagnosing individual samples; SE = number of true positive samples/(number of true positive samples + number of false negative samples)
- SP: specificity of the test for diagnosing individual samples; SP = number of true negative samples/(number of true negative samples + number of false positive samples).
- SE_P: sensitivity of the test for diagnosing sample pools; SE_P = number of true positive pools/(number of true positive pools + number of false negative pools)
- SP_P: specificity of the test for diagnosing sample pools; SP_P = number of true negative pools/(number of true negative pools + number of false positive pools) and
- P_A: apparent prevalence of the pools; P_A = number of test-positive pools/total number of pools.

Mathematical basis of pool testing

Each sample that is part of a pool can either be truly positive or truly negative. We can therefore represent the infection status of each individual sample using an independent and identically distributed (i.i.d.) Bernoulli(p) random variable x_i, such that:

$$x_i = \begin{cases} 1 \text{ with probability } p \\ 0 \text{ with probability } (1-p) \end{cases} \quad (1)$$

The probability P_T of finding at least one positive sample in a pool of size k is given by Chiang & Reeves (1962):

$$P_T = 1 - (1-p)^k \quad (2)$$

Therefore, the result of the pool test may be represented by an i.i.d. Bernoulli(P_T) random variable y_j where:

$$y_j = \begin{cases} 1 \text{ with probability } 1-(1-p)^k \\ 0 \text{ with probability } (1-p)^k \end{cases} \quad (3)$$

By grouping the individual samples in n pools of size k_j ($j = 1,2,\ldots,n$), the apparent pool prevalence P_A may be calculated by dividing the number of positive pools by the number of pools tested:

$$P_A = \frac{\sum_{j=1}^{n} y_j}{n} \quad (4)$$

When the diagnostic test is perfect (i.e. $SE = SP = 1$), the apparent pool prevalence P_A is an unbiased estimator of the true probability P_T of obtaining at least one positive sample in the group of samples. In case of constant pool sizes (i.e. $k_j = k$), the true individual-based prevalence p may then be estimated by transforming Eqn (2):

$$p \doteq 1 - (1 - P_A)^{\frac{1}{k}} \tag{5}$$

In practice, however, perfect diagnostic tests do not exist (i.e. SE and/or $SP \neq 1$). As the probability SE (SP) of a truly positive (negative) individual sample to test positive (negative) is less than 100%, the probability SE_P (SP_P) of a truly positive (negative) pool of samples to test positive (negative) will equally be less than 100% (Munoz-Zanzi *et al.*, 2006). The relationship between the individual and pool test characteristics are given by Boelaert *et al.* (2000):

$$SE_P = 1 - [(1 - SE)^{kp} * SP^{k(1-p)}] \tag{6}$$

$$SP_P = SP^k \tag{7}$$

Under the condition of an imperfect diagnostic test, P_A no longer is an unbiased estimator of P_T, as the apparent prevalence P_A then estimates a function of the pool sensitivity SE_P and the pool specificity SP_P:

$$P_A \approx P_T * SE_P + (1 - P_T)(1 - SP_P) \tag{8}$$

By replacing P_T by its expression in Eqn (2), we obtain:

$$P_A \approx [1 - (1 - p)^k] * SE_P + (1 - p)^k(1 - SP_P) \tag{9}$$

From Eqn (9), and considering a fixed pool size, it can be shown that:

$$p \doteq 1 - \left[\frac{SE_P - P_A}{SE_P + SP_P - 1}\right]^{1/k} \tag{10}$$

Prevalence estimation models

Estimating infection prevalence in vector populations based on pooled samples can be done using either a frequentist or a Bayesian statistical philosophy (Messam *et al.*, 2008). In the frequentist philosophy, a population parameter is assumed to have a single fixed value, and a 95% 'confidence' interval for this parameter is considered to be the interval which should contain the real value in 95% of the cases, if the test were to be repeated a large number of times. Methods known as being frequentist typically do not take account of existing (*a priori*) knowledge on the diagnostic test characteristics (but see Cowling *et al.*, 1999; Vansteelandt *et al.*, 2000 for examples of using known fixed sensitivity and specificity when estimating the prevalence). In the present study, we will examine the frequentist maximum likelihood (ML) estimation model that assumes perfect tests (Farrington, 1992; Cowling *et al.*, 1999; Williams & Moffitt, 2001, 2005).

In the Bayesian philosophy, population parameters are assumed to have an intrinsic probability distribution. This approach produces a 95% 'credibility' interval, meaning that, given the observations, there is a 95% chance that the true value lies between the limits of the credibility interval (Gardner, 2002). The Bayesian philosophy combines field data and expert (*a priori*) opinions in a single model, combining test results with *a priori* information on the test characteristics (i.e. sensitivity and specificity), resulting in an *a posteriori* probability distribution of the prevalence. The Bayesian approach can incorporate uncertainty on unknown and variable parameters (such as diagnostic test characteristics), and allows for a flexible combination of complex equations. These methods, initially developed for estimating the prevalence based on individual samples (Enoe *et al.*, 2000), were thereafter extended to estimation based on the pooled samples (Mendoza-Blanco *et al.*, 1996; Cowling *et al.*, 1999). In the present study, two categories of Bayesian approaches will be appraised, i.e. the Monte Carlo Markov-Chain (MCMC) Bayesian models and the Exact Bayesian Computation (EBC) models.

The 14 models presented and compared in the present study, are categorized in 3 groups (i.e. frequentist, MCMC and EBC models). The models are implemented in R (R Development Core Team, 2010) and OpenBUGS (Lunn *et al.*, 2009), which are both freely available and open-source software environments. R is an interactive programming language developed for statistical analyses and graphics, whereas OpenBUGS is a package specifically designed to conduct Bayesian analyses. An overview of the models and their implementation methods is given in Table 1. The source code of these models is presented in Appendix S1.

Frequentist models. An estimate of the true prevalence can be computed using the ML approach (Farrington, 1992; Cowling *et al.*, 1999; Williams & Moffitt, 2001, 2005), the method implemented in currently available software programs such as PoolScreen™ (Katholi *et al.*, 1995) and PooledInfRate (Biggerstaff, 2009). For fixed pool sizes and under the assumption of a perfect test, the ML estimator of the prevalence can be calculated according to Eqn (5). In case of a fixed pool size and a test with known test characteristics, the ML estimate corresponds to Eqn (10). Uncertainty regarding the test characteristics is, however, typically not incorporated in frequentist models of disease prevalence.

In case of variable pool sizes, the ML estimator is more difficult to express in a closed form. One possible approach is to use a generalized linear model (GLM) (Farrington, 1992), i.e. a logistic regression equation with the complementary log-log (*cloglog*) link function. For estimating the ML of the prevalence, the *cloglog* function can be written as follows: $cloglog(P_T) = \ln[-\ln(1-P_T)]$, with P_T the true prevalence of the pools. Considering Eqn (2), we obtain the following:

$$\begin{aligned} cloglog(P_T) &= \ln[-\ln[(1-p)^k]] \\ &= \ln[-k * \ln(1-p)] \\ &= \ln(k) + \ln[-\ln(1-p)] \end{aligned} \tag{11}$$

Table 1. Overview of the 14 models used to estimate the individual-based prevalence based on diagnostic test results obtained through testing pools of samples.

Group	Case			Model	Implementation
Frequentist models	Perfect individual test	Wald-type confidence interval		1-WALD	R
		Inverted likelihood-ratio test based confidence interval		1-INV	R
Monte Carlo Markov-Chain (MCMC) Bayesian models	Perfect individual test	With regression	*logit* link	2aR-LOG	OpenBUGS
			cloglog link	2aR-CLL	OpenBUGS
		Without regression		2aD	OpenBUGS
	Imperfect individual test, constant test characteristics	With regression	*logit* link	2bR-LOG	OpenBUGS
			cloglog link	2bR-CLL	OpenBUGS
		Without regression		2bD	OpenBUGS
	Imperfect individual test, test characteristics dependent on pool weights	With regression	*logit* link	2cR-LOG	OpenBUGS
			cloglog link	2cR-CLL	OpenBUGS
		Without regression		2cD	OpenBUGS
Exact Bayesian Computation (EBC) models	Perfect individual test			3a	R/C++
	Imperfect individual test, constant test characteristics			3b	R/C++
	Imperfect individual test, test characteristics dependent on pool weights			3c	R/C++

Model annotations: 1 = Frequentist *cloglog* model with Wald-type confidence intervals (WALD) or confidence intervals calculated by inverting the likelihood-ratio test (INV); 2 = Bayesian MCMC models based on regression using a *cloglog* link (R-CLL), regression using a *logit* link (R-LOG), or direct estimation (D); 3 = Exact Bayesian computation model. The letter 'a' in the model annotations indicates models considering perfect test characteristics, the letter 'b' indicates models considering imperfect test characteristics, and the letter 'c' indicates models considering imperfect test characteristics that are a function of the pool weights.

There are two components in the right side of this equation, i.e. $\ln(k)$, the so-called offset, and $\ln[-\ln(1-p)]$, which is a constant (β). We can therefore write:

$$cloglog(P_T) = \ln(k) + \beta \tag{12}$$

and the individual-based prevalence p can then be obtained by back-transforming β:

$$p = 1 - \exp[-\exp(\beta)] \tag{13}$$

The R code is shown in the Appendix S1 for estimation of the *cloglog* model parameters. The presented R code creates Wald-type confidence intervals for the prevalence [of the form estimate ± multiplier (standard error)], as well as confidence intervals constructed by inverting the likelihood-ratio test (Williams & Moffitt, 2001). Both methods are shown in the results, and denoted as Model 1-WALD and Model 1-INV, respectively.

Monte Carlo Markov-Chain Bayesian models. The MCMC Bayesian models (hereafter just called MCMC models) apply an iterative Bayesian technique for estimating the prevalence. Unlike the frequentist ML estimation methods, the MCMC models use a given prior probability distribution Prob(θ) of the population parameter θ, which is combined with the observed likelihood Prob(Data|θ) to obtain the posterior distribution Prob(θ|Data):

$$\text{Prob}(\theta|\text{Data}) \propto \text{Prob}(\theta) * \text{Prob}(\text{Data}|\theta) \tag{14}$$

This allows for the incorporation of prior knowledge on the parameters to be estimated, such as prior knowledge on the uncertainty of the test characteristics in the prevalence estimation.

According to the test characteristics, three cases are considered:

Model 2a: Perfect individual test ($SE = SP = 1$).

Model 2b: Imperfect individual test (SE and/or $SP \neq 1$), test characteristics independent of the pool weight. The pool sensitivity and specificity are derived directly from Eqns (6) and (7).

Model 2c: Imperfect individual test (SE and/or $SP \neq 1$), test sensitivity as a function of the pool weight. In order to explore the effect of modulating factors, the pool sensitivity is formulated as a decreasing logistic function of the pool weight Eqn (15). The inversely proportional relation between the sensitivity and the pool weight corresponds to the assumption that the inhibition of the PCR test may be proportional to the quantity of blood (and inhibitors), and thus to the weight of the pools of vectors. This assumption is then incorporated in the models to correct the pool sensitivity given by Eqn (6).

$$wSE_P = \left[\frac{\exp(a + b * \text{Weight})}{1 + \exp(a + b * \text{Weight})}\right] \tag{15}$$

Where,

wSE_P = weight-adjusted pool sensitivity;
a = constant or intercept, reflecting the pool sensitivity when the weight is not taken into account;

b = regression coefficient or slope, linking the sensitivity and the sample weight; and
Weight = weight of the pool in grams.

The parameter a is given by Eqn (16), with SE_p the unadjusted (and thus maximum) test sensitivity:

$$a = \ln\left[\frac{SE_p}{1 - SE_p}\right] \tag{16}$$

For each of these three cases, two types of MCMC models are presented, a type that uses regression models to relate the observed pool results to the individual-based prevalence (Models 2aR, 2bR and 2cR), and a type that does not use regression models but directly estimates the individual-based prevalence (Models 2aD, 2bD and 2cD):

Models 2aR, 2bR and 2cR: MCMC regression models

As in the frequentist approach, a GLM for binomial data may also be used for Bayesian prevalence estimations. These models use logistic regression equations with the *cloglog* or the *logit* link to relate the pool prevalence to individual-based prevalence (Farrington, 1992; Vansteelandt *et al.*, 2000) (referred to as CLL and LOG in the model designations). The *cloglog* function was previously described (see the frequentist model). The *logit* function is given by: $\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \Phi$, where Φ is a constant. By back-transforming Φ, the individual-based prevalence p may be obtained:

$$p = \frac{\exp(\Phi)}{1 + \exp(\Phi)} \tag{17}$$

Models 2aD, 2bD, and 2cD: MCMC models without regression

Different from the models with regression using a (*logit* or *cloglog*) link function, the models without regression use Eqns (5) and (10) and thus directly estimate the prevalence from the apparent prevalence (hence the notation 'D').

Exact Bayesian Computation (EBC) models. Unlike the frequentist and MCMC models, the EBC models are based on a likelihood-free methodology. The EBC models repeatedly simulate a random population parameter based on the parameter's prior distribution, and use this value to simulate a data set. Through an accept/reject algorithm, the parameter values that exactly result in the observed data set are accepted, whereas other values are rejected. The distribution of the accepted values is then regarded as a sample from the population parameter's posterior distribution. This approach is similar to the Approximate Bayesian Computation (ABC) algorithms (Pritchard *et al.*, 1999; Tanaka *et al.*, 2006; Sisson *et al.*, 2007), except that the EBC algorithm requires an exact match to the data and hence is not approximate. The EBC approach is very flexible and powerful and can be used for very complex models, but may become computationally intensive. The EBC models developed for the present study are implemented in the R software, using code that calls on additional functions written in C++ (for increased speed). As for the MCMC models, three cases are considered (Model 3a, Model 3b and Model 3c).

Model 3a: Perfect individual test: the following steps are performed:

- sample a random prevalence p^* based on *a prior* Uniform (0, 1) distribution;
- use this prevalence to generate a vector of Bernoulli(p^*) random variables x_i^*;
- assign these values to pools according to the sampling design used in the experiment;
- determine the pool results y_j^*: if a pool contains at least one positive sample, it is positive ($y_i^* = 1$), otherwise it is negative ($y_j^* = 0$);
- accept p^* if it resulted in the observed pool results (i.e., $y^* = y$), otherwise reject it;
- repeat these steps until the required number of accepted values is obtained; and
- compute the accepted prevalence average and 95% credibility interval.

Model 3b: Imperfect individual test, constant test characteristics

The procedure of this model is similar to the one described for Model 3a, but now imperfect test sensitivity and specificity are included. Based on, for example, expert opinion, probability distributions for the test characteristics are defined. Experts can for example define the minimum and maximum values of a parameter, which can be used to create the appropriate distribution. These distributions are then used to draw random sensitivities and specificities. Based on these randomly drawn values, Bernoulli random variables are generated and used to determine the apparent pool results.

Model 3c: Imperfect individual test, test sensitivity as a function of the pool weight

In addition to Model 3b, this model includes a probability distribution for the slope parameter b of the test sensitivity modulating function, as explained for Model 2c. Randomly generated values for b are then included in Eqn (15), in order to obtain weight-adjusted pool sensitivities. These values are then used to calculate the apparent pool results as in Model 3b.

Data used to test the models

The 14 statistical models are tested on different datasets. Two types of datasets will be used: simulated data and data derived from the literature.

Simulated data. To examine the efficiency of the different methods in estimating the individual-based prevalence, we simulate the process of pooled testing based on a predefined and thus known individual-based prevalence (Gu *et al.*, 2004). Furthermore, published studies generally do not report on the possible pool-related or external factors that can influence test characteristics. Only simulated data allow exploring the strength of the models to deal with these effects.

Three cases of increasing complexity with respect to the test characteristics have been considered in the previous section. In agreement with the characteristics used in these cases,

Table 2. Simulated sample pool results, based on two times six pools of 5, 10, 20, 30, 40 and 50 samples.

Prevalence (%)	Case a (SE = SP = 1; b = 0)	Case b (SE = 0.8; SP = 1; b = 0)	Case c (SE = 0.8; SP = 1; b = −0.07)
0	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
1	0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
4	0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1	0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1	0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0
10	1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1	0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0

The pool weights used for case c were 5.59, 21.69, 21.52, 39.95, 58.39, 79.88, 8.51, 11.90, 27.86, 40.00, 86.52 and 105.45. 0's represent negative pools, 1's represent positive pools.

pool data were generated. For the case presuming a perfect individual test (case a), the individual test sensitivity and specificity were fixed to 100%. For the cases presuming an imperfect individual test (cases b and c), an individual sensitivity of the PCR was set to be 80%. The individual test specificity was assumed to be 100%. In addition, for the case where test sensitivity was presumed to be imperfect and a function of the pool weight (case c), the slope parameter b of the logistic regression function was set to -0.07. These choices were made arbitrarily, but represent realistic scenarios of the respective parameters.

For each case, four simulations were generated, based on four typically low prevalence values (i.e. 0%, 1%, 4% and 10%). The pooling strategy was the same in each simulation, i.e. two times six pools of 5, 10, 20, 30, 40 and 50 samples. This strategy is based on Gu *et al.* (2004), who concluded that variable pool sizes allow more accurate prevalence estimations than constant pool sizes. Indeed the pooling strategy has an important influence on the prevalence estimation. For details on how to design an optimal pooling strategy, we refer to the work of Gu and colleagues (Gu *et al.*, 2004, 2008; Gu & Novak, 2004). It is beyond the scope of the present study to compare different pooling strategies, hence the choice for a single strategy which is proven to be reliable under a wide range of prevalences (Gu *et al.*, 2004).

As we focus on comparing the different estimation models, we selected the most representative pool results for the respective case and prevalence. For each combination, 5000 simulations were generated and the frequency of each different combination of pool results was calculated. The combination that corresponded to the median frequency was selected, thereby avoiding the selection of extreme pool results.

The simulation program used for this purpose is written in R, and is presented in the Appendix S1. It generates pool results under the three cases considered in the simulation study:

The simulation program runs through the following steps to generate the data:

- generate 310 Bernoulli(p) random variables (1 for each individual sample in the 12 pools), with p the selected individual-based prevalence;
- generate 310 sample weights, by exponentiating random variables from the standard normal distribution;
- regroup both results in 2 × 6 pools of size 5, 10, 20, 30, 40 and 50;
- determine the true pool results y_{Tj}: if a pool contains at least one positive sample, it is positive ($y_{Tj} = 1$), otherwise it is negative ($y_{Tj} = 0$);
- calculate SE_P and SP_P, based on Eqns (6) and (7), using the user-defined values for SE and SP;
- model the sensitivity as a logistic function of the pool weights previously calculated (based on a Eqns (15) and (16), and a user-defined slope b);
- generate a Bernoulli(SE_P) and Bernoulli(SP_P) random variable; and
- determine the apparent pool results y_{Aj}: if $y_{Tj} = 1$ and $SE_P = 0$, let $y_{Aj} = 0$ (false negative), if $y_{Tj} = 0$ and $SP_P = 0$, let $y_{Aj} = 1$ (false positive), otherwise let $y_{Aj} = y_{Tj}$.

The results of the simulations used in the present study are presented in Table 2.

Literature data. In addition to the simulated data, published data were used to assess the models in realistic (thus imperfect) settings. The selected data are from studies of vectors transmitting parasites to humans and animals, i.e. sandflies (infected with *Leishmania* spp.) and ticks [*Amblyomma americanum* (Linnaeus) (Acari: Ixodidae), the lone star tick, infected with *Borrelia lonestari* (Barbour) (Spirochaetales: Spirochaetaceae)]. Table 3 summarizes the selected data. More datasets were tested (e.g. Williams & Moffitt, 2001) but patterns were similar to those reported here.

Application of the different models

The simulated and published pool results were analysed with the 14 different models. For the models considering an imperfect test sensitivity, we used a uniform prior distribution for sensitivity ranging from 60% to 95%. For the models that consider test sensitivity as a function of the pool weight, we applied a Normal($\mu = -0.07$, $\sigma^2 = 0.04$) prior for the slope parameter, based on Gelman *et al.* (2008). Finally, a Uniform(0,1) prior was used for the prevalence in all MCMC models [note that a Uniform(0,1) distribution corresponds to a Beta(1,1) distribution, as applied in our code]. The choice of these distributions and their values reflects expert opinion on the uncertain parameters. Depending on the situation, users applying these models should select the appropriate prior distributions and parameters.

The nine MCMC models are implemented in OpenBUGS, using 10 000 iterations and two chains. Only 8000 iterations were retained, because the first 2000 iterations correspond to the 'burn-in' iterations needed to obtain a stable and

Table 3. Literature data used to test the different estimation models.

Species, location and year of collection	Pool sizes*	Number of pools	Number of positive pools	Source
Sandflies, Alfacar, Spain, 1998	30	10	2	1
Sandflies, Viznar, Spain, 2004	30	17	0	1
Sandflies, Duhabi, Nepal, 2007	5, 10, 11, 11, 9, **11***, 13, **2***, 17, 8, 13, 16, 16, 17, 1, 2, 3, 12, 11, 20, 20, 20, 20, 19, 1, 20, 20, 20, 18, 2, 5, 14, 14, 14, 14, 4, 14, 18, 19, 6, 20, 14, 1, 9, 20, 4, 1, 1, 1, 16, 14, 2, 8, 6	54	2	2
Sandflies, Aurabani, Nepal, 2007	2, 1, 6, 10, 1, 7, 1, 4, **1***, 3	10	1	2
Ticks (*Amblyomma americanum*), Pulaski County (Missouri), U.S.A.	5	7	2	3

1 = Martin-Sanchez *et al*. (2006); 2 = Bhattarai *et al.* (2009); 3 = Bacon *et al.* (2005).
*In case of variable pool sizes, the positive pools are marked with an asterisk.

converging distribution of the parameters to estimate. The Brooks–Gelman–Rubin (BGR) statistic was used to check the convergence of the models in accordance with the principles of Brooks & Gelman (1998). For the 3 EBC models, 500 accepted simulations were retained.

Results

Estimates and confidence or credibility intervals obtained with the different models are presented in Tables 4–7. For all MCMC models in OpenBUGS, the BGR statistic indicated convergence of the various models.

Simulation study

The first simulation assumed a perfect diagnostic test sensitivity and specificity. All models that considered a perfect individual test were able to produce estimates that match the true prevalence reasonably well. The estimates of the Bayesian models that wrongly considered an imperfect individual test were biased upwards. Confidence and credible intervals widened as the true prevalence increased.

The frequentist *cloglog* model estimates were lower than those obtained using the other models. The results of the three MCMC models belonging to case a were similar, as was the case for the results of the three MCMC models belonging to case b. The results and credibility intervals of the EBC models approximated those of the corresponding MCMC models.

When all pools had the same results (i.e. if all pools were positive or negative), the Wald-type confidence interval for the frequentist *cloglog* model resulted in a degenerative 0–100 confidence interval. The likelihood-ratio-based method, on the other hand, did permit calculation of an appropriate confidence interval.

For the second simulated case, the test sensitivity was set to be imperfect, but independent of the pool weight. The estimates of the frequentist model and those of the Bayesian models assuming perfect test characteristics (case a) were lower than the true prevalence. The estimates of the Bayesian models that considered an imperfect individual test (case b) were closer to the true prevalence.

For this final simulation, with test sensitivity assumed to be imperfect and inversely related to the pool weight, all models were fitted to the simulated pools data. The four Bayesian models belonging to case c generated similar results, which tended to be better estimators of the true prevalence than the models belonging to case a and b. Especially for a high true prevalence, the models that did not incorporate the extra uncertainty on the slope parameter were not able to generate confidence or credibility intervals that included the true prevalence. However, the credibility intervals of the models belonging to case c were wider than those of the other models. It can also be noted that it was only possible to obtain a prevalence close to the one used during the generation of the data if the variance of the prior (i.e. constraint) on the slope parameter of the logistic function linking the sensitivity and the weight was very small.

Literature data

With the literature data, similar trends were observed as obtained with the simulation study, such as the lower estimations obtained with the frequentist *cloglog* model and the similar results obtained with the corresponding MCMC and EBC models. The absolute difference between the frequentist model, the Bayesian models assuming perfect tests and the Bayesian models presuming imperfect tests increased as the apparent prevalence increased.

Discussion

With the appearance of Bluetongue in Europe (Purse *et al.*, 2005) and the continuous presence of important vector-borne diseases such as Malaria and Leishmaniasis in various parts of the world, arthropod vectors remain a world concern and deserve the required attention. The estimation of the prevalence of pathogens within these vector populations will remain a crucial prerequisite in understanding the economic and health threat and burden arising from these diseases (Inci *et al.*,

Table 4. Prevalence estimation results and corresponding 95% confidence or credibility intervals for sample pool results simulated under the assumption of a perfect individual test (case a).

	True prevalence (%)			
Model	0	1	4	10
1-WALD	0.00 (0.00–100.00)	1.16 (0.37–3.61)	3.51 (1.56–7.82)	16.69 (6.16–40.83)
1-INV	0.00 (0.00–0.62)	1.16 (0.29–3.02)	3.51 (1.34–7.61)	16.69 (6.47–38.21)
2aR-CLL	0.33 (0.00–0.96)	1.55 (0.43–3.40)	4.26 (1.64–8.25)	21.06 (7.94–41.70)
2aR-LOG	0.33 (0.00–0.96)	1.55 (0.43–3.40)	4.26 (1.64–8.25)	21.06 (7.94–41.70)
2aD	0.33 (0.00–0.96)	1.55 (0.43–3.40)	4.26 (1.64–8.25)	21.06 (7.94–41.70)
2bR-CLL	0.59 (0.00–1.61)	2.49 (1.04–4.64)	5.67 (2.53–10.39)	25.64 (10.54–50.20)
2bR-LOG	0.59 (0.00–1.61)	2.49 (1.04–4.64)	5.67 (2.53–10.39)	25.64 (10.54–50.20)
2bD	0.59 (0.00–1.61)	2.49 (1.04–4.64)	5.67 (2.53–10.39)	25.64 (10.54–50.20)
3a	0.31 (0.00–0.90)	1.56 (0.46–3.60)	4.27 (1.43–8.67)	20.83 (8.60–42.66)
3b	0.62 (0.00–1.57)	2.44 (0.91–4.48)	5.52 (2.55–10.10)	26.06 (11.68–49.52)

Model annotations: 1 = Frequentist *cloglog* model with Wald-type confidence intervals (WALD) or confidence intervals calculated by inverting the likelihood-ratio test (INV); 2 = Bayesian MCMC models based on regression using a *cloglog* link (R-CLL), regression using a *logit* link (R-LOG), or direct estimation (D); 3 = Exact Bayesian computation model. The letter 'a' in the model annotations indicates models considering perfect test characteristics, while the letter 'b' indicates models considering imperfect test characteristics.

Table 5. Prevalence estimation results and corresponding 95% confidence or credibility intervals for sample pool results simulated under the assumption of an imperfect individual test with constant test characteristics (case b).

	True prevalence (%)			
Model	0	1	4	10
1-WALD	0.00 (0.00–100.00)	0.34 (0.05–2.41)	2.57 (1.07–6.07)	6.21 (2.83–13.33)
1-INV	0.00 (0.00–0.62)	0.34 (0.02–1.51)	2.57 (0.90–5.74)	6.21 (2.59–13.40)
2aR-CLL	0.33 (0.00–0.96)	0.69 (0.09–1.94)	3.17 (1.14–6.31)	7.70 (3.03–15.00)
2aR-LOG	0.33 (0.00–0.96)	0.69 (0.09–1.94)	3.17 (1.14–6.31)	7.70 (3.03–15.00)
2aD	0.33 (0.00–0.96)	0.69 (0.09–1.94)	3.17 (1.14–6.31)	7.70 (3.03–15.00)
2bR-CLL	0.59 (0.00–1.61)	1.28 (0.31–2.82)	4.41 (2.01–7.95)	9.57 (4.39–17.74)
2bR-LOG	0.59 (0.00–1.61)	1.28 (0.31–2.82)	4.41 (2.01–7.95)	9.57 (4.39–17.74)
2bD	0.59 (0.00–1.61)	1.28 (0.31–2.82)	4.41 (2.01–7.95)	9.57 (4.39–17.74)
3a	0.31 (0.00–0.90)	0.69 (0.09–1.87)	3.07 (0.95–6.42)	7.78 (3.14–14.19)
3b	0.62 (0.00–1.57)	1.32 (0.30–3.03)	4.35 (1.94–7.94)	9.44 (4.39–16.71)

Model annotations: 1 = Frequentist *cloglog* model with Wald-type confidence intervals (WALD) or confidence intervals calculated by inverting the likelihood-ratio test (INV); 2 = Bayesian MCMC models based on regression using a *cloglog* link (R-CLL), regression using a *logit* link (R-LOG), or direct estimation (D); 3 = Exact Bayesian computation model. The letter 'a' in the model annotations indicates models considering perfect test characteristics, while the letter 'b' indicates models considering imperfect test characteristics.

2007). Owing to various logistic and financial reasons, the testing of pools of individual vector samples is often preferred (Munoz-Zanzi *et al.*, 2006). As the underlying prevalence in the vector population cannot be readily estimated from these pools, adapted analytical techniques are warranted. The present study reviewed frequentist and Bayesian methods for estimating individual-based prevalence using results from pools of samples. In addition, a code was developed to estimate the prevalence based on EBC models, a flexible and easily extendable method which has seldom been used in the past (Bhattarai *et al.*, 2009). The code of all models used in the present study is included in Appendix S1, allowing users to easily apply and analyse the presented models.

To evaluate the efficiency of the various models, pool results were generated based on known true prevalences. Three situations were tested. First, under the assumption of a 'reference standard' test (i.e. perfect test sensitivity and specificity), all models that assumed perfect individual tests were able to reliably reproduce the true prevalence. If the test characteristics were wrongly assumed to be suboptimal, however, the results tended to be biased, as expected. In our case, the wrongly assumed suboptimal sensitivity tended to overestimate the true prevalence, as compensation for the wrongly assumed high proportion of false negatives. Confidence and credibility intervals widened as the true prevalence increased, which might be a consequence of the increasing variability as the true prevalence approaches 50%. In addition, as the individual-based prevalence increases, the chance of a pool being positive, thus obtaining a high pool prevalence, increases. In this situation, models may fail to generate precise and unbiased estimates. Gu *et al.* (2004, 2008) showed that the variable size pooling design used in the present study provides better estimates than a constant size pooling design, and this for a range of true infection rates between 0% and 10%. The upper limit of this range refers to the situation where the likelihood that all pools will be positive becomes too high, and may explain the

Table 6. Prevalence estimation results and corresponding 95% confidence or credibility intervals for sample pool results simulated under the assumption of an imperfect individual test with test sensitivity as a function of the pool weight (case c).

	True Prevalence (%)			
Model	0	1	4	10
1-WALD	0.00 (0.00–100.00)	0.00 (0.00–100.00)	0.33 (0.05–2.40)	3.78 (1.69–8.32)
1-INV	0.00 (0.00–0.62)	0.00 (0.00–0.62)	0.33 (0.02–1.46)	3.78 (1.58–7.60)
2aR-CLL	0.33 (0.00–0.96)	0.33 (0.00–0.96)	0.67 (0.09–1.86)	4.39 (1.84–8.16)
2aR-LOG	0.33 (0.00–0.96)	0.33 (0.00–0.96)	0.67 (0.09–1.86)	4.39 (1.84–8.16)
2aD	0.33 (0.00–0.96)	0.33 (0.00–0.96)	0.67 (0.09–1.86)	4.39 (1.84–8.16)
2bR-CLL	0.59 (0.00–1.61)	0.59 (0.00–1.61)	1.29 (0.31–2.83)	5.83 (2.97–9.97)
2bR-LOG	0.59 (0.00–1.61)	0.59 (0.00–1.61)	1.29 (0.31–2.83)	5.83 (2.97–9.97)
2bD	0.59 (0.00–1.61)	0.59 (0.00–1.61)	1.29 (0.31–2.83)	5.83 (2.97–9.97)
2cR-CLL	8.93 (0.00–26.22)	8.93 (0.00–26.22)	11.26 (0.76–32.53)	12.66 (2.66–32.25)
2cR-LOG	8.93 (0.00–26.22)	8.93 (0.00–26.22)	11.26 (0.76–32.53)	12.66 (2.66–32.25)
2cD	8.93 (0.00–26.22)	8.93 (0.00–26.22)	11.26 (0.76–32.53)	12.66 (2.66–32.25)
3a	0.31 (0.00–0.93)	0.33 (0.00–0.98)	0.68 (0.10–2.04)	4.39 (2.01–7.60)
3b	0.61 (0.00–1.56)	0.58 (0.00–1.53)	1.29 (0.30–2.81)	5.82 (2.94–9.90)
3c	9.74 (0.00–26.92)	9.27 (0.00–28.72)	11.02 (0.88–31.40)	12.44 (2.84–32.37)

Model annotations: 1 = Frequentist *cloglog* model with Wald-type confidence intervals (WALD) or confidence intervals calculated by inverting the likelihood-ratio test (INV); 2 = Bayesian MCMC models based on regression using a *cloglog* link (R-CLL), regression using a *logit* link (R-LOG), or direct estimation (D); 3 = Exact Bayesian computation model. The letter 'a' in the model annotations indicates models considering perfect test characteristics, the letter 'b' indicates models considering imperfect test characteristics, and the letter 'c' indicates models considering imperfect test characteristics that are a function of the pool weights.

Table 7. Prevalence estimation results and corresponding 95% confidence or credibility intervals for previously reported sample pool results.

	Data source				
Model	Viznar	Duhabi	Alfacar	Aurabani	Pulaski County
1-WALD	0.00 (0.00–100.00)	0.33 (0.08–1.32)	0.74 (0.19–2.94)	2.78 (0.37–19.33)	6.51 (1.66–23.73)
1-INV	0.00 (0.00–0.38)	0.33 (0.05–1.01)	0.74 (0.12–2.28)	2.78 (0.16–11.67)	6.51 (1.11–18.93)
2aR-CLL	0.20 (0.00–0.58)	0.49 (0.10–1.18)	1.11 (0.23–2.67)	5.26 (0.68–14.11)	9.20 (1.99–21.08)
2aR-LOG	0.20 (0.00–0.58)	0.49 (0.10–1.18)	1.11 (0.23–2.67)	5.26 (0.68–14.11)	9.20 (1.99–21.08)
2aD	0.20 (0.00–0.58)	0.49 (0.10–1.18)	1.11 (0.23–2.67)	5.26 (0.68–14.11)	9.20 (1.99–21.08)
2bR-CLL	0.45 (0.00–1.18)	1.56 (0.57–2.97)	1.95 (0.71–3.82)	9.88 (2.50–21.40)	14.42 (5.31–28.14)
2bR-LOG	0.45 (0.00–1.18)	1.56 (0.57–2.97)	1.95 (0.71–3.82)	9.88 (2.50–21.40)	14.42 (5.31–28.14)
2bD	0.45 (0.00–1.18)	1.56 (0.57–2.97)	1.95 (0.71–3.82)	9.88 (2.50–21.40)	14.42 (5.31–28.14)
3a	0.19 (0.00–0.56)	0.47 (0.10–1.25)	1.12 (0.21–2.51)	5.35 (0.49–13.63)	9.00 (2.31–21.63)
3b	0.47 (0.00–1.19)	1.55 (0.68–2.84)	1.97 (0.68–3.99)	10.02 (2.31–19.00)	14.67 (5.56–29.16)

Model annotations: 1 = Frequentist *cloglog* model with Wald-type confidence intervals (WALD) or confidence intervals calculated by inverting the likelihood-ratio test (INV); 2 = Bayesian MCMC models based on regression using a *cloglog* link (R-CLL), regression using a *logit* link (R-LOG), or direct estimation (D); 3 = Exact Bayesian computation model. The letter 'a' in the model annotations indicates models considering perfect test characteristics, while the letter 'b' indicates models considering imperfect test characteristics.

overestimation and uncertainty for our simulation based on a 10% true individual-based prevalence.

In real settings, test characteristics are rarely perfect, and false-negative or false-positive results are known to occur in all kinds of diagnostic tests, even in PCR assays, which are generally assumed to have superior test characteristics (Katholi *et al.*, 1995). The second simulation therefore generated pool results based on imperfect individual test sensitivity. The pool sensitivity was then calculated based on the pool size and expected prevalence (Boelaert *et al.*, 2000). In contrast to the previous case, models assuming perfect test characteristics, most notably the frequentist *cloglog* model, were not able to reproduce the true prevalence, but tended to underestimate it. This is logical as not accounting for a suboptimal sensitivity will result in false negatives lowering the apparent prevalence. Bayesian models that took in account prior knowledge on the test sensitivity proved to be more reliable. These results confirm the importance of incorporating prior knowledge on test characteristics (Williams & Moffitt, 2010). This prior knowledge is generally derived through expert elicitation, a process in which one or more experts express their opinion on the value or range of a certain parameter. Different expert elicitation methods exist, ranging from simple face-to-face interviews to complex structured processes involving several experts (see also O'Hagan *et al.* (2006) for a guide on expert elicitation).

Although our second simulation considered an imperfect, but constant individual test sensitivity, it has been reported that endogenous (pool-related) and exogenous (external) factors can potentially influence the characteristics of diagnostic

tests, so these no longer can be considered to be constant. Examples of endogenous factors are the proportion of females in the pool (as certain female vectors can engorge themselves with considerable amounts of blood) and the presence and concentration of inhibitors (Schwartz *et al.*, 1997; Munoz-Zanzi *et al.*, 2006). Exogenous factors may relate to field-related factors such as the locality where and the season when the samples were taken, or to laboratory-related factors such as the experience of the lab technician (Speybroeck *et al.*, 2011). Therefore, one of the objectives of the present study was to investigate the possibility of taking such factors into account in the analysis of sample pools. In a third simulation, we incorporated in the data generation process the inhibition effect on the pool sensitivity, considered to be proportional to the weight of the pools. Especially for high true prevalences, models that were able to take in account this effect of pool weight produced more accurate prevalence estimates than the models that only took into account an imperfect test sensitivity. However, these more complex models generated relatively large and asymmetric credibility intervals, reflecting the additional level of uncertainty. Furthermore, the present study revealed the need for very accurate and precise prior knowledge (i.e. very narrow priors) on the parameters that relate the sensitivity to the influencing factor (in our case, the slope parameter b that related the sensitivity to the pool weight). Only when using the same slope parameter b as the one used in the data generation for creating a narrow prior, the results fitted. Such precise information is, unfortunately, not available under real-life conditions. This observation brings a paradox to the surface: if no reference standard test exists, how can the required knowledge on the test characteristics be obtained, and how can the field data be reliably corrected? Where possible, researchers should therefore try to avoid this situation by collecting and testing samples that do not contain such modulating factors. For example, this could mean excluding blood-fed specimens from the pools to be tested to avoid the inhibitory blood effect on PCRs. Pathogens will develop and amplify inside the body of competent vectors, thus detection is not limited to those arthropods that have recently blood-fed. Fed specimens are usually very few in captures (unless specific traps for blood-fed arthropods are used) and are easily recognized while preparing the pool. However, avoiding modulating factors, may, unfortunately, not be feasible for all possible modulators. Whether or not modulating factors are taken in account, the estimated prevalence will be strongly influenced by what the experts say. Indeed, the expression of two different opinions by two experts on the value and range of a parameter will result in two different estimations of the individual-based prevalence. Scientists should therefore be very transparent in reporting their results. The expert opinions, i.e. the distributions used, need to be presented explicitly, as well as the reasons for these opinions. In addition, the plausibility of the elicited parameters and the consequent results should always be debated in the light of the biological understanding of the vector and pathogen. The use of several diagnostic tests on the same pools may reduce the dependency on expert opinions, but combining diagnostic test information on pools of samples is a yet unexplored research field. A reasonable option may therefore be to report results under different scenarios of test characteristics, and to be fully transparent about the applied test characteristics.

Conclusion

Prior knowledge on the characteristics of the applied diagnostic test is essential for generating unbiased or best-guess prevalence estimates. Therefore, researchers should have a good understanding of the test characteristics under ideal conditions. In addition, the identification of modulating factors is crucial, and carefully designed experiments should be conducted to generate as accurate information on these factors as possible. Finally, it is clear that only a Bayesian context will give acceptable results as it is the only approach that properly includes known uncertainty on diagnostic test characteristics.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: DOI: 10.1111/j.1365-2915.2012.01015.x

Appendix S1. R, OpenBUGS and C++ Code.

References

Abel, U., Schosser, R. & Suss, J. (1999) Estimating the prevalence of infectious agents using pooled samples: biometrical considerations. *Zentralblatt fur Bakteriologie : International Journal of Medical Microbiology*, **289**, 550–563.

Al-Soud, W.A., Ouis, I.S., Li, D.Q., Ljungh, S. & Wadstrom, T. (2005) Characterization of the PCR inhibitory effect of bile to optimize real-time PCR detection of *Helicobacter* species. *FEMS Immunology and Medical Microbiology*, **44**, 177–182.

Bacon, R.M., Pilgard, M.A., Johnson, B.J., Piesman, J., Biggerstaff, B.J. & Quintana, M. (2005) Rapid detection methods and prevalence estimation for *Borrelia lonestari* glpQ in *Amblyomma americanum* (Acari: Ixodidae) pools of unequal size. *Vector Borne and Zoonotic Diseases*, **5**, 146–156.

Berkvens, D., Speybroeck, N., Praet, N., Adel, A. & Lesaffre, E. (2006) Estimating disease prevalence in a Bayesian framework using probabilistic constraints. *Epidemiology*, **17**, 145–153.

Bhattarai, N.R., Das, M.L., Rijal, S. *et al.* (2009) Natural infection of *Phlebotomus argentipes* with *Leishmania* and other trypanosomatids in a visceral leishmaniasis endemic region of Nepal. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **103**, 1087–1092.

Biggerstaff, B.J. (2009) *PooledInfRate, Version 4.0: A Microsoft® Office Excel© Add-In to Compute Prevalence Estimates from Pooled Samples.* Centers for Disease Control and Prevention, Fort Collins, CO.

Boelaert, F., Walravens, K., Biront, P., Vermeersch, J.P., Berkvens, D. & Godfroid, J. (2000) Prevalence of paratuberculosis (Johne's disease) in the Belgian cattle population. *Veterinary Microbiology*, **77**, 269–281.

Brooks, S.P. & Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–455.

Chiang, C.L. & Reeves, W.C. (1962) Statistical estimation of virus infection rates in mosquito vector populations. *American Journal of Hygiene*, **75**, 377–391.

Cowling, D.W., Gardner, I.A. & Johnson, W.O. (1999) Comparison of methods for estimation of individual-level prevalence based on pooled samples. *Preventive Veterinary Medicine*, **39**, 211–225.

Dorfman, R. (1943) The detection of defective members of large populations. *Annals of Mathematical Statistics*, **14**, 436–440.

Enoe, C., Georgiadis, M.P. & Johnson, W.O. (2000) Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive Veterinary Medicine*, **45**, 61–81.

Farrington, C.P. (1992) Estimating prevalence by group testing using generalized linear models. *Statistics in Medicine*, **11**, 1591–1597.

Gardner, I.A. (2002) The utility of Bayes' theorem and Bayesian inference in veterinary clinical practice and research. *Australian Veterinary Journal*, **80**, 758–761.

Gelman, A., Jakulin, A., Pittau, M.G. & Su, Y.S. (2008) A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, **2**, 1360–1383.

Gratz, N.G. (1999) Emerging and resurging vector-borne diseases. *Annual Review of Entomology*, **44**, 51–75.

Gu, W. & Novak, R.J. (2004) Short report: detection probability of arbovirus infection in mosquito populations. *The American Journal of Tropical Medicine and Hygiene*, **71**, 636–638.

Gu, W., Lampman, R. & Novak, R.J. (2004) Assessment of arbovirus vector infection rates using variable size pooling. *Medical and Veterinary Entomology*, **18**, 200–204.

Gu, W., Unnasch, T.R., Katholi, C.R., Lampman, R. & Novak, R.J. (2008) Fundamental issues in mosquito surveillance for arboviral transmission. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **102**, 817–822.

Inci, A., Ica, A., Yildirim, A. *et al.* (2007) Economical impact of tropical theileriosis in the Cappadocia region of Turkey. *Parasitology Research*, **101** (Suppl. 2), S171–S174.

Katholi, C.R., Toe, L., Merriweather, A. & Unnasch, T.R. (1995) Determining the prevalence of *Onchocerca volvulus* infection in vector populations by polymerase chain reaction screening of pools of black flies. *The Journal of Infectious Diseases*, **172**, 1414–1417.

Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009) The BUGS project: evolution, critique and future directions. *Statistics in Medicine*, **28**, 3049–3067.

Martin-Sanchez, J., Gallego, M., Baron, S., Castillejo, S. & Morillas-Marquez, F. (2006) Pool screen PCR for estimating the prevalence of *Leishmania infantum* infection in sandflies (Diptera: Nematocera, Phlebotomidae). *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **100**, 527–532.

Mendoza-Blanco, J.R., Tu, X.M. & Iyengar, S. (1996) Bayesian inference on prevalence using a missing-data approach with simulation-based techniques: applications to HIV screening. *Statistics in Medicine*, **15**, 2161–2176.

Messam, L.L.M., Branscum, A.J., Collins, M.T. & Gardner, I.A. (2008) Frequentist and Bayesian approaches to prevalence estimation using examples from Johne's disease. *Animal Health Research Reviews/Conference of Research Workers in Animal Diseases*, **9**, 1–23.

Munoz-Zanzi, C., Thurmond, M., Hietala, S. & Johnson, W. (2006) Factors affecting sensitivity and specificity of pooled-sample testing for diagnosis of low prevalence infections. *Preventive Veterinary Medicine*, **74**, 309–322.

O'Hagan, A., Buck, C.E., Daneshkhah, A. *et al.* (2006) *Uncertain Judgments: Eliciting Expert Probabilities.* John Wiley & Sons Ltd, Chichester.

Oshaghi, M.A., Rasolian, M., Shirzadi, M.R., Mohtarami, F. & Doosti, S. (2010) First report on isolation of *Leishmania tropica* from sandflies of a classical urban Cutaneous leishmaniasis focus in southern Iran. *Experimental Parasitology*, **126**, 445–450.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. & Feldman, M.W. (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, **16**, 1791–1798.

Purse, B.V., Mellor, P.S., Rogers, D.J., Samuel, A.R., Mertens, P.P. & Baylis, M. (2005) Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology*, **3**, 171–181.

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.

Schwartz, I., Varde, S., Nadelman, R.B., Wormser, G.P. & Fish, D. (1997) Inhibition of efficient polymerase chain reaction amplification of *Borrelia burgdorferi* DNA in blood-fed ticks. *The American Journal of Tropical Medicine and Hygiene*, **56**, 339–342.

Sisson, S.A., Fan, Y. & Tanaka, M.M. (2007) Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 1760–1765.

Speybroeck, N., Praet, N., Claes, F. *et al.* (2011) True versus apparent malaria infection prevalence: the contribution of a Bayesian approach. *PLos ONE*, **6**, e16705.

Sutherst, R.W., Ingram, J.S. & Scherm, H. (1998) Global change and vector-borne diseases. *Parasitology Today*, **14**, 297–299.

Tanaka, M.M., Francis, A.R., Luciani, F. & Sisson, S.A. (2006) Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics*, **173**, 1511–1520.

Vansteelandt, S., Goetghebeur, E. & Verstraeten, T. (2000) Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics*, **56**, 1126–1133.

Watt, D., Sparagano, O., Brown, C.G. & Walker, A.R. (1997) Use of the polymerase chain reaction for identification and quantification of *Theileria parva* protozoa in *Rhipicephalus appendiculatus* ticks. *Parasitology Research*, **83**, 359–363.

Williams, C.J. & 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.

Williams, C.J. & Moffitt, C.M. (2005) Estimation of pathogen prevalence in pooled samples using maximum likelihood methods and open-source software. *Journal of Aquatic Animal Health*, **18**, 149–155.

Williams, C.J. & Moffitt, C.M. (2010) Estimation of fish and wildlife disease prevalence from imperfect diagnostic tests on pooled samples with varying pool sizes. *Ecological Informatics*, **5**, 273–280.

Yameogo, L., Toe, L., Hougard, J.M., Boatin, B.A. & Unnasch, T.R. (1999) Pool screen polymerase chain reaction for estimating the prevalence of *Onchocerca volvulus* infection in *Simulium damnosum* sensu lato: results of a field trial in an area subject to successful vector control. *The American Journal of Tropical Medicine and Hygiene*, **60**, 124–128.

Accepted 15 January 2012
First published online 8 April 2012