ESTIMATING DISEASE PREVALENCE IN A BAYESIAN FRAMEWORK USING PROBABILISTIC CONSTRAINTS

D. Berkvens[1], N. Speybroeck[1], N. Praet[1], A. Adel[2] & E. Lesaffre[3]

Epidemiology 17, 145-53

March 27, 2006

[1] Institute of Tropical Medicine, Department of Animal Health, Antwerp, Belgium

[2] Ecole Nationale Vétèrinaire, Département Clinique, El Harach, Alger, Algeria

[3] Biostatistical Centre, Catholic University of Leuven, Leuven, Belgium

Contents

Abstract

Estimating the prevalence of a disease from the results of one or more diagnostic tests applied to individuals of unknown disease status invariably means that, in the absence of a gold standard and without external constraints, more parameters must be estimated than the data permit. Two assumptions regularly made in the literature, namely that the test characteristics (sensitivity and specificity) are constant over populations and/or the tests are conditionally independent given the true disease status, have been criticized recently in the literature as being unrealistic[1]. Nevertheless, to estimate the prevalence restrictions on the parameter estimates need to be imposed. We consider two types of restrictions: deterministic and probabilistic restrictions, the latter arising in a Bayesian framework when expert knowledge is available. Further, we consider two possible parameterisations allowing incorporation of these restrictions. The first is an extension of the approach of Gardner et al.[2] and Dendukuri & Joseph[3], assuming conditional dependence, to more than two diagnostic tests. It is argued here that this system of equations is difficult to combine with expert opinions. The second approach, based on conditional probabilities, looks more promising to us and is developed in a Bayesian context. To evaluate the combination of data with the (deterministic and probabilistic) constraints the recently developed DIC and p_D are applied[4], together with an appropriate Bayesian P-value. The latter approach is illustrated using data collected in a study on the prevalence of porcine cysticercosis and is verified using external data.

Keywords: Bayesian inference, diagnostic tests, deterministic constraints, probabilistic constraints, prevalence of disease

1 Introduction

Diagnostic tests form an essential part of all disciplines of epidemiology, because an estimate of the true prevalence of the disease, infection or condition is required in a number of situations.

Suppose that D^+ (D^-) indicates that a subject is diseased (disease-free) and T^+ (T^-) indicates a positive (negative) result on a diagnostic test T. In the presence of a gold standard, the number of diseased subjects (n_{D^+}) and disease-free subjects (n_{D^-}) are known (Table 1). A gold standard can be a diagnostic test with both test sensitivity and test specificity equal to one or (e.g.) an experiment where a proportion of the subjects are artificially infected. The columns Diseased and Disease-free in Table 1 represent this situation and constitute the so-called full table, i.e. the table where the distinction between the two infection status categories can be made.

	Diseased	Disease-free	Total
+ test result	$n_{T^+\|D^+}$	$n_{T^+\|D^-}$	n_{T^+}
- test result	$n_{T^-\|D^+}$	$n_{T^-\|D^-}$	n_{T^-}
Total	n_{D^+}	n_{D^-}	n

Table 1: Two-by-two contingency table when testing n subjects for disease D with one diagnostic test T. D^+ (D^-) indicates that the subject is (not) diseased, T^+ (T^-) indicates a positive (negative) result with test T.

From Table 1, sensitivity (Se) and specificity (Sp) of the test are estimable by $n_{T^-|D^-}/n_{D^-}$ and $n_{T^-|D^-}/n_{D^-}$, respectively. On the other hand, in a field observation only the probability of a positive test result can be directly estimated, i.e. $P\left(T^+\right) = n_{T^+}/n$ (the apparent prevalence). The column Total in Table 1 is actually the marginal or collapsed table over the diseased and disease-free subjects and represents this situation. When Se and Sp are known, the true prevalence $P(D^+)$ can be estimated using the expression[5]

$$P\left(D^+\right) = \frac{P\left(T^+\right) + Sp - 1}{Se + Sp - 1} \tag{1}$$

Unfortunately, Se and Sp are rarely known exactly and must be estimated from data. Hence we need to take into account the sampling

variability with which the prevalence is estimated, which could be done using the approach of Rogan & Gladen[5].

Some traditional textbooks on diagnostic testing still refer to the test sensitivity and specificity as values, which are intrinsic to the diagnostic test, i.e. constant and universally applicable[6,7]. Our own experience indicates that both test sensitivity and specificity vary with external factors[8,9,10] and others[1,11]. Consequently, test sensitivity and specificity, as traditionally defined, are purely theoretical concepts determined in the population used to validate the test. Therefore, when using a diagnostic test in the population of interest, the characteristics of that population must be used to get an improved estimate of Se and Sp[1]. Observe, though, that constancy of Se and Sp over different populations is still being made[12].

For a long time it was assumed that two (or more) diagnostic tests are conditionally independent on the disease status, i.e. $P\left(T_1^+ \cap T_2^+ | D^+\right) = P\left(T_1^+ | D^+\right) P\left(T_2^+ | D^+\right)$ [13,14,15] and similar for other possible test results and for disease-free subjects. When the two diagnostic tests have a similar biological basis, as is often the case, the conditional independence assumption is untenable. Toft et al.[16] review the possible pitfalls when using the Hui-Walter paradigm in real life, particularly the problems encountered when trying to stratify the population into two or more sub-populations with different true prevalence but constant test characteristics.

When the above two simplifying assumptions cannot be made, estimation of the true prevalence becomes either impossible or extra information must be added to the estimation process. Indeed, applying h tests to each individual implies that 2^{h+1} - 1 parameters must be estimated, being the true prevalence, h test sensitivities and specificities plus $2\sum_{i=2}^{h}\left(\begin{array}{c} h \\ i \end{array}\right) = 2\left(2^h - h - 1\right) = 2^{h+1} - 2h - 2$ parameters describing the dependence of the h tests, given the true disease status of the subject. Yet only 2^h - 1 parameters can be estimated, because only data from the collapsed table (over disease status) are available. Consequently, the true prevalence of the disease cannot be estimated if no constraints are imposed on the parameters. The most popular constraint has been to assume conditional independence.

Table 2 shows the maximum number of parameters that can be estimated and the number of parameters that need to be estimated as

a function of the number of diagnostic tests, as well as the number of parameters to be estimated given conditional independence of the tests.

Number of tests	Maximum number of estimable parameters	Parameters to be estimated under conditional dependence	Parameters to be estimated under conditional independence
1	1	3	3
2	3	7	5
3	7	15	7
4	15	31	9
5	31	63	11
h	2^h-1	2^{h+1}-1	2h+1

Table 2: Maximum number of estimable parameters, number of parameters to be estimated in the absence of conditional independence and under conditional independence as a function of the number of tests per subject.

In particular, Table 2 indicates that, under conditional independence, parameters can be estimated for $h \geq 3$ and that for $h \geq 4$ the number of estimable parameters actually exceeds the number of parameters to estimate.

Estimating the true prevalence thus becomes a matter of adding constraints on the parameters. These constraints must come from external sources, e.g. previous similar studies, expert opinion, etc. Hence, the estimated true prevalence and test characteristics will be the result of a combination of the data (test results) and the external information on these test characteristics, which is the best that can be obtained. Consequently, several authors have suggested a Bayesian approach to incorporate this external information by specifying prior distributions on the parameters obtained from eliciting the opinion of experts [17,18]. Most often, it is attempted to incorporate prior knowledge on sensitivity and specificity. Unfortunately, in practice experts often do not have and cannot have (cfr. non-constant sensitivity and specificity) a clear-cut opinion on these test characteristics. As a result the experts opinions will regularly be in conflict with the actually observed data. Of course, the Bayesian framework allows making the

prior distributions more diffuse but this will, in our context, often render the parameters inestimable. In this paper we will show that, if possible, prior information on conditional probabilities is easier to specify.

To verify whether the prior information is in conflict with the test results the recently developed deviance information criterion (DIC)[4] and an appropriate Bayesian P-value can be used[19]. To quantify the impact of the constraints, the effective number of estimated parameters (pD) of the model can be calculated[4].

In Section 2, two parameterisations to model conditional dependence are discussed. In Section 3 we distinguish between deterministic and probabilistic constraints and show that p_D can be used to quantify the effect of these constraints on the number of effectively estimated parameters. In the next section (Section 4)we indicate that DIC and an appropriate Bayesian P-value can pinpoint a conflict between the prior information and the test results. In Section 5 we examine the behaviour of DIC, p_D and the Bayesian P-value, using a theoretical data set. In Section 6 we apply the second model developed in Section 2 to field data. A discussion of our approach and the results follows in the last section (Section 7).

Markov Chain Monte Carlo (MCMC) estimations were carried out in WinBUGS 1.4 (http://www.mrc-bsu.cam.ac.uk/bugs/[20]. Additional calculations were performed in R (http://www.r-project.org/), making extensive use of the bugs() function (http://www.stat.columbia.edu/∼gelman/bugsR/[19]. The software developed for the evaluation of DIC, p_D and the Bayesian P-value can be found in Appendix A.

2 Modelling conditional dependence between tests through conditional probabilities

Gardner at al.[2] and Dendukuri & Joseph[3] calculated the probabilities of the different outcomes as a function of test sensitivities, test specificities and covariances, when two diagnostic tests are applied to all subjects. Further, these authors also suggest combining prior information on these parameters with the test results in a Bayesian manner. Their results can be expanded to more than two tests (Appendix B)[21]. However, the prior distributions for the covariances (i.e. generalised beta distributions) are quite difficult to elicit from

experts, because they cannot be related to real-life situations. Although not well recognized in the literature, we argue that this is equally true for the sensitivity parameters, the reason being that the sensitivity of a diagnostic test needs to be determined in experimental conditions (and hence also quite distinct from real-life settings) on a small number of subjects. In contrast, the specificity of a test can be determined somewhat easier in a population, which is known to be disease-free.

Eliciting information from experts on the conditional performance of one test given the results of another test could be much easier in certain cases. For instance, we argue that a question like *"What is the probability that a subject tests positively in test2, given that it is diseased and that it has tested positively in test1?"* (relating the characteristics of two tests applied on the same subject) can be easier to answer because the experts usually have one or more so called reference tests (very often with a very high specificity) and know the performance of other tests in relation to the reference test in the infected and uninfected sub-populations. Model (2), given by

$$P\left(T_1^{i_1} \cap \cdots \cap T_h^{i_h}\right) = P\left(D^+\right) \prod_{t=1}^{h} \left[(1 - i_t) - (-1)^{i_t} P\left(T_t^+ \mid D^+ \bigcap_{t'=1|t>1}^{t-1} T_{t'}^{i_{t'}}\right)\right] + \left[1 - P\left(D^+\right)\right] \prod_{t=1}^{h} \left[i_t + (-1)^{i_t} P\left(T_t^- \mid D^- \bigcap_{t'=1|t>1}^{t-1} T_{t'}^{i_{t'}}\right)\right] \tag{2}$$

expresses the cell probabilities of the collapsed $2^{(h+1)}$ table (hence of a 2^h table) in terms of the prevalence of the disease, the sensitivity and specificity of the first test and conditional probabilities. In Table 3 the different conditional probabilities are listed in a hierarchical fashion: parameters θ_1-θ_3 are used when only a single test is applied, θ_1-θ_7 are used for two tests, θ_1-θ_{15} for three tests and θ_1-θ_{31} for four tests. In Table 4 expressions are given to calculate the prevalence and the test characteristics from the parameters defined in Table 3. Finally, in Table 5 the equations are given to calculate the cell probabilities of the different test result combinations when h = 4. When fewer than four tests are used, the probabilities can be extracted from these equations, by dropping excess terms: e.g. P(111) = $\theta_1\theta_2\theta_4\theta_8$ + (1-θ_1)(1-θ_3)(1-θ_7)(1-θ_{15}).

prevalence	$P(D^+)$	θ_1
Se_1	$P\left(T_1^+ \mid D^+\right)$	θ_2
Sp_1	$P\left(T_1^- \mid D^-\right)$	θ_3
	$P\left(T_2^+ \mid D^+ \cap T_1^+\right)$	θ_4
	$P\left(T_2^+ \mid D^+ \cap T_1^-\right)$	θ_5
	$P\left(T_2^- \mid D^- \cap T_1^-\right)$	θ_6
	$P\left(T_2^- \mid D^- \cap T_1^+\right)$	θ_7
	$P\left(T_3^+ \mid D^+ \cap T_1^+ \cap T_2^+\right)$	θ_8
	$P\left(T_3^+ \mid D^+ \cap T_1^+ \cap T_2^-\right)$	θ_9
	$P\left(T_3^+ \mid D^+ \cap T_1^- \cap T_2^+\right)$	θ_{10}
	$P\left(T_3^+ \mid D^+ \cap T_1^- \cap T_2^-\right)$	θ_{11}
	$P\left(T_3^- \mid D^- \cap T_1^- \cap T_2^-\right)$	θ_{12}
	$P\left(T_3^- \mid D^- \cap T_1^- \cap T_2^+\right)$	θ_{13}
	$P\left(T_3^- \mid D^- \cap T_1^+ \cap T_2^-\right)$	θ_{14}
	$P\left(T_3^- \mid D^- \cap T_1^+ \cap T_2^+\right)$	θ_{15}
	$P\left(T_4^+ \mid D^+ \cap T_1^+ \cap T_2^+ \cap T_3^+\right)$	θ_{16}
	$P\left(T_4^+ \mid D^+ \cap T_1^+ \cap T_2^+ \cap T_3^-\right)$	θ_{17}
	$P\left(T_4^+ \mid D^+ \cap T_1^+ \cap T_2^- \cap T_3^+\right)$	θ_{18}
	$P\left(T_4^+ \mid D^+ \cap T_1^+ \cap T_2^- \cap T_3^-\right)$	θ_{19}
	$P\left(T_4^+ \mid D^+ \cap T_1^- \cap T_2^+ \cap T_3^+\right)$	θ_{20}
	$P\left(T_4^+ \mid D^+ \cap T_1^- \cap T_2^+ \cap T_3^-\right)$	θ_{21}
	$P\left(T_4^+ \mid D^+ \cap T_1^- \cap T_2^- \cap T_3^+\right)$	θ_{22}
	$P\left(T_4^+ \mid D^+ \cap T_1^- \cap T_2^- \cap T_3^-\right)$	θ_{23}
	$P\left(T_4^- \mid D^- \cap T_1^- \cap T_2^- \cap T_3^-\right)$	θ_{24}
	$P\left(T_4^- \mid D^- \cap T_1^- \cap T_2^- \cap T_3^+\right)$	θ_{25}
	$P\left(T_4^- \mid D^- \cap T_1^- \cap T_2^+ \cap T_3^-\right)$	θ_{26}
	$P\left(T_4^- \mid D^- \cap T_1^- \cap T_2^+ \cap T_3^+\right)$	θ_{27}
	$P\left(T_4^- \mid D^- \cap T_1^+ \cap T_2^- \cap T_3^-\right)$	θ_{28}
	$P\left(T_4^- \mid D^- \cap T_1^+ \cap T_2^- \cap T_3^+\right)$	θ_{29}
	$P\left(T_4^- \mid D^- \cap T_1^+ \cap T_2^+ \cap T_3^-\right)$	θ_{30}
	$P\left(T_4^- \mid D^- \cap T_1^+ \cap T_2^+ \cap T_3^+\right)$	θ_{31}

Table 3: Conditional probabilities

p	$=$	θ_1
Se_1	$=$	θ_2
Sp_1	$=$	θ_3
Se_2	$=$	$\theta_2\theta_4 + (1-\theta_2)\,\theta_5$
Sp_2	$=$	$\theta_3\theta_6 + (1-\theta_3)\,\theta_7$
Se_3	$=$	$\theta_2\left\{\theta_4\theta_8 + (1-\theta_4)\,\theta_9\right\} + (1-\theta_2)\left\{\theta_5\theta_{10} + (1-\theta_5)\,\theta_{11}\right\}$
Sp_3	$=$	$\theta_3\left\{\theta_6\theta_{12} + (1-\theta_6)\,\theta_{13}\right\} + (1-\theta_3)\left\{\theta_7\theta_{14} + (1-\theta_7)\,\theta_{15}\right\}$
Se_4	$=$	$\theta_2\left\{\theta_4\left\langle\theta_8\theta_{16} + (1-\theta_8)\,\theta_{17}\right\rangle + (1-\theta_4)\left\langle\theta_9\theta_{18} + (1-\theta_9)\,\theta_{19}\right\rangle\right\}$ $+ (1-\theta_2)\left\{\theta_5\left\langle\theta_{10}\theta_{20} + (1-\theta_{10})\,\theta_{21}\right\rangle + (1-\theta_5)\left\langle\theta_{11}\theta_{22} + (1-\theta_{11})\,\theta_{23}\right\rangle\right\}$
Sp_4	$=$	$\theta_3\left\{\theta_6\left\langle\theta_{12}\theta_{24} + (1-\theta_{12})\,\theta_{25}\right\rangle + (1-\theta_6)\left\langle\theta_{13}\theta_{26} + (1-\theta_{13})\,\theta_{27}\right\rangle\right\}$ $+ (1-\theta_3)\left\{\theta_7\left\langle\theta_{14}\theta_{28} + (1-\theta_{14})\,\theta_{29}\right\rangle + (1-\theta_7)\left\langle\theta_{15}\theta_{30} + (1-\theta_{15})\,\theta_{31}\right\rangle\right\}$

Table 4: Parameters

$P(0000)^{\text{a}}$	$=$	$\theta_1(1-\theta_2)(1-\theta_5)(1-\theta_{11})(1-\theta_{23}) + (1-\theta_1)\theta_3\theta_6\theta_{12}\theta_{24}$
$P(0001)$	$=$	$\theta_1(1-\theta_2)(1-\theta_5)(1-\theta_{11})\theta_{23} + (1-\theta_1)\theta_3\theta_6\theta_{12}(1-\theta_{24})$
$P(0010)$	$=$	$\theta_1(1-\theta_2)(1-\theta_5)\theta_{11}(1-\theta_{22}) + (1-\theta_1)\theta_3\theta_6(1-\theta_{12})\theta_{25}$
$P(0011)$	$=$	$\theta_1(1-\theta_2)(1-\theta_5)\theta_{11}\theta_{22} + (1-\theta_1)\theta_3\theta_6(1-\theta_{12})(1-\theta_{25})$
$P(0100)$	$=$	$\theta_1(1-\theta_2)\theta_5(1-\theta_{10})(1-\theta_{21}) + (1-\theta_1)\theta_3(1-\theta_6)\theta_{13}\theta_{26}$
$P(0101)$	$=$	$\theta_1(1-\theta_2)\theta_5(1-\theta_{10})\theta_{21} + (1-\theta_1)\theta_3(1-\theta_6)\theta_{13}(1-\theta_{26})$
$P(0110)$	$=$	$\theta_1(1-\theta_2)\theta_5\theta_{10}(1-\theta_{20}) + (1-\theta_1)\theta_3(1-\theta_6)(1-\theta_{13})\theta_{27}$
$P(0111)$	$=$	$\theta_1(1-\theta_2)\theta_5\theta_{10}\theta_{20} + (1-\theta_1)\theta_3(1-\theta_6)(1-\theta_{13})(1-\theta_{27})$
$P(1000)$	$=$	$\theta_1\theta_2(1-\theta_4)(1-\theta_9)(1-\theta_{19}) + (1-\theta_1)(1-\theta_3)\theta_7\theta_{14}\theta_{28}$
$P(1001)$	$=$	$\theta_1\theta_2(1-\theta_4)(1-\theta_9)\theta_{19} + (1-\theta_1)(1-\theta_3)\theta_7\theta_{14}(1-\theta_{28})$
$P(1010)$	$=$	$\theta_1\theta_2(1-\theta_4)\theta_9(1-\theta_{18}) + (1-\theta_1)(1-\theta_3)\theta_7(1-\theta_{14})\theta_{29}$
$P(1011)$	$=$	$\theta_1\theta_2(1-\theta_4)\theta_9\theta_{18} + (1-\theta_1)(1-\theta_3)\theta_7(1-\theta_{14})(1-\theta_{29})$
$P(1100)$	$=$	$\theta_1\theta_2\theta_4(1-\theta_8)(1-\theta_{17}) + (1-\theta_1)(1-\theta_3)(1-\theta_7)\theta_{15}\theta_{30}$
$P(1101)$	$=$	$\theta_1\theta_2\theta_4(1-\theta_8)\theta_{17} + (1-\theta_1)(1-\theta_3)(1-\theta_7)\theta_{15}(1-\theta_{30})$
$P(1110)$	$=$	$\theta_1\theta_2\theta_4\theta_8(1-\theta_{16}) + (1-\theta_1)(1-\theta_3)(1-\theta_7)(1-\theta_{15})\theta_{31}$
$P(1111)^{\text{b}}$	$=$	$\theta_1\theta_2\theta_4\theta_8\theta_{16} + (1-\theta_1)(1-\theta_3)(1-\theta_7)(1-\theta_{15})(1-\theta_{31})$

[a] $P(0000) = P\left(T_1^- \cap T_2^- \cap T_3^- \cap T_4^-\right)$
[b] $P(1111) = P\left(T_1^+ \cap T_2^+ \cap T_3^+ \cap T_4^+\right)$

Table 5: Test result probabilities

3 Deterministic versus probabilistic constraints and the use of p_D

Constraints on the parameters need to be imposed to estimate the prevalence and the test characteristics using (2). We classify these constraints into two types: deterministic and probabilistic. Setting Se (or Sp) to a particular value is an example of a deterministic constraint, as is also assuming conditional independence. Specifying a prior distribution for a parameter or for a function of parameters (like a contrast) is an example of a probabilistic constraint in a Bayesian setting.

In a frequentist context, m independent deterministic constraints reduce the number of parameters to estimate exactly by m. For instance, when h = 2 assuming conditional independence involves 2 independent constraints reducing the number of parameters to estimate from 7 to 5 (see Table 2). When fixing the specificity of one test to, say 1, the number of parameters to estimate is further reduced by one. In a Bayesian context, things are more difficult since it is not immediately clear what the impact is of a probabilistic constraint on the number of parameters to estimate. In this context, Spiegelhalter et al.[4] proposed to measure the effective number of estimated parameters in a fitted statistical model by p_D. This measure is not an integer anymore, even for a deterministic constraint, as it is calculated as the difference of the posterior mean of the deviance and the deviance evaluated in the posterior mean. More details are given in next section.

4 Measuring the discordance of the prior information with the observed test results

4.1 Introduction

As described in the introduction, experts have great difficulties to express their prior knowledge in quantitative terms (cfr. sensitivity and specificity). Our experience shows that often the prior information is in conflict with the actual observed data. In the context of diagnostic testing this is evidently a crucial handicap. In the statistical literature, several authors have addressed this problem[22], but it is not

immediately clear how the proposed measures for discordance can be implemented in our context. Here, two measures will be proposed. The first one is based on a Bayesian goodness-of-fit test leading to a Bayesian P-value. The second one uses the recently introduced deviance information criterion (DIC)[4]. Both measures will be reviewed here in the context of analysing collapsed tables of diagnostic test data in a Bayesian manner. Although not absolutely necessary, we will assume that Bayesian estimation is done via MCMC sampling and reference will be made to the WinBUGS software.

4.2 Computation of Bayesian P-value and the deviance information criterion

4.2.1 Multinomial model

Assume that there are h diagnostic tests implying a $2^{(h+1)}$ (=k) contingency table applied to N subjects, then it can often be assumed that the cell frequencies r_i (i=1, ..., k) follow a multinomial distribution, namely:

$$r_i \sim Mult\left(\pi_1, \ldots, \pi_k; N\right) \qquad (i = 1, \ldots, k)$$

The cell probabilities π_i (i=1, ..., k) can depend on q parameters θ_j (j=1, ..., q), i.e. $\pi_i = \pi_i\left(\theta_1, \ldots, \theta_q\right)$. The parameters are classically estimated by maximum likelihood yielding the maximum likelihood estimate (MLE) of θ_j denoted as $\hat{\theta}_j$, which gives also the MLE of π_i denoted as $\hat{\pi}_i$. Typically $q < (k-1)$, when $q = (k-1)$ then the model is called saturated and if the π-parameters are 1-1 to the θ-parameters, then the MLE of π_i is r_i/N (i=1, ..., k). When $q > (k-1)$ the model is called "overspecified" such that the parameters are actually unidentifiable. In this paper we are dealing with a 2^h contingency table, hence k = 2^h - 1 but the parameters are the cell probabilities belonging to the $2^{(h+1)}$- contingency table implying q = $2^{(h+1)} - 1 > k - 1$. The same is true, say for h = 4, when the parameters are θ_1 to θ_{31} of Table 4. Hence, we are dealing in this paper with an overspecified multinomial model. In general, the fit of the estimated multinomial model to the data can be measured by the deviance[23] given by $dev = 2\left[\sum_{i=1}^{k} r_i \ln\left(\frac{r_i}{N}\right) - \sum_{i=1}^{k} r_i \ln\left(\hat{\pi}_i\right)\right]$, whereby

the first term in the expression represents the perfect fit of the model to the data.

4.2.2 Bayesian P-value

It is illustrative to see what the deviance would be if we sampled (pseudo-)data under the actually fitted model, i.e. if we randomly generated observations using the current model. If the observed data are obtained from the fitted model and we would repeatedly sample such pseudo observations, then the average of $\tilde{\delta} = \text{dev}_o\text{-dev}_r$, the difference of the deviance of the sampled pseudo-data (dev_r) with the deviance of the observed data (dev_o) would be about zero. A positive value for $\tilde{\delta}$ indicates that the fit of the model to the observed data is worse than the fit of the model to the pseudo-data generated under that same model.

In a MCMC context (at convergence) $\tilde{\delta}$ can be calculated at each iteration based the current sampled value of the parameter estimates θ_j (j=1, ..., q) yielding a chain $\tilde{\delta}_1, \ldots, \tilde{\delta}_T$. The average $\frac{1}{T}\sum_{t=1}^{T} I\left(\tilde{\delta}_t > 0\right)$, whereby I(x) =1 when $x \geq 0$ and 0 otherwise, yields the posterior estimate of $P(\tilde{\delta} > 0)$. The difference $\tilde{\delta}$ is called a posterior predictive check[19]. (The estimate of) $P(\tilde{\delta} > 0)$ is called a Bayesian P-value and expresses the fit of the assumed model to the data, i.e. when small the assumed model is probably not appropriate for the data at hand.

4.2.3 Deviance Information Criterion

Akaikes Information Criterion (AIC) for a multinomial model is given by

$$AIC = -2\sum_{i=1}^{k} r_i \ln\left(\hat{\pi}_i\right) + 2\,p$$

where p is the the number of parameters that are estimated. A low value for AIC indicates a good fit of the model to the data after having penalized for the number of parameters that are estimated. Hence, a complicated model (high value of p) needs to give a real improvement of the fit to result in a lower AIC. In this way AIC can be used to select models. The deviance information criterion

(DIC)[4] is a generalisation of AIC to a Bayesian setting where p is replaced by the Bayesian equivalent, namely p_D and $\hat{\pi}_i$ in the first term is replaced by a Bayesian estimate (e.g. posterior mean). Since $p_D = -2\sum_{i=1}^{k} r_i \ln(\pi_i) + 2\sum_{i=1}^{k} r_i \ln(\hat{\pi}_i)^4$, where the first part of the expression is the posterior mean of (2 x minus) log-likelihood, DIC can be rewritten as $DIC = -4\sum_{i=1}^{k} r_i \ln(\pi_i) + 2\sum_{i=1}^{k} r_i \ln(\hat{\pi}_i)$. A model with a small DIC has the same property as AIC but in a Bayesian context. However, it is important to stress that a model is now the combination of the data (likelihood) and the prior information (prior distribution) yielding a posterior distribution. Thus a correct likelihood with a wrong prior distribution will yield bad predictions and this will be indicated by a high value for DIC. Observe that, if an MCMC approach is used, DIC should be calculated after convergence. A condition for DIC to be a reliable measure as well as for p_D is that the likelihood is log-concave in its parameters[4,20] when for the Bayesian estimate the posterior mean is chosen. Further, the posterior mean needs to be a good summary measure for the parameters (i.e. the posterior distributions need to be fairly symmetric). When log-concavity is not satisfied, p_D can become negative and also the estimate for DIC is not trustworthy.

4.2.4 Comparison between the measures

The principal difference between the above defined Bayesian P-value (depending on $\tilde{\delta}$) and DIC is that the deviance information criterion penalises for the number of parameters required to be estimated in the model.

5 Behaviour of DIC, p_D and Bayesian P-value

5.1 DIC and p_D

In this section, we will discuss the performance of DIC and p_D in the context of a possibly overspecified multinomial model. That is, we look at the behaviour of DIC and p_D when $q > (k-1)$ and we will focus on model (2). When $q \leq (k-1)$ we expect $p_D \approx k - 1 -$

q. Unfortunately, this will not necessarily be the case for model (2) as this model is not log-concave in its parameters. Things become worse when $q > (k-1)$ because then the log-likelihood must be flat around the maximum likelihood estimate if no constraints have been imposed. However, if the multinomial model is parameterised in its multinomial probabilities, i.e. in π_i (i=1, , k-1), then for all cases the log-likelihood will be concave in its parameters. Consequently, we suggest evaluating DIC and p_D always in the posterior mean of π_i (i=1, , k-1). But, there is one remaining problem, though, namely that p_D (if based on the multinomial probabilities) is always smaller than k - 1 whether or not the model has been overspecified. To have an idea when the model has been overspecified we suggest calculating p_D also using the posterior means of its parameters, i.e. for model (2) on the posterior means of the parameters θ_1 to θ_{31} for h = 4. Empirical evidence shows that without sufficient constraints in that case p_D is negative resulting in a diagnostic that can indicate whether all our parameters are estimable or not.

To exemplify our reasoning in the previous paragraph we take the case of h = 1, hence when there is only one diagnostic test and the multinomial model contains only two cells, i.e. $\pi_1 = P\left(T_1^+\right)$ and $\pi_2 = P\left(T_1^-\right)$. In this case, $\pi_1 = \theta_1\theta_2 + (1-\theta_1)(1-\theta_3)$ and $\pi_2 = \theta_1(1-\theta_2) + (1-\theta_1)\theta_3$. The log-likelihood is not concave in θ_1, θ_2 and θ_3 but clearly it is in π_1 (we can neglect π_2 since it is $(1-\pi_1)$. Without any constraints on θ_1, θ_2 and θ_3 the multinomial parameter π_1 will vary freely, thus $p_D \approx 1$ if based on the posterior mean of π_1. But, experience showed that p_D becomes negative when based on θ_1, θ_2 and θ_3. When putting constraints on θ_1, θ_2 and θ_3 nothing will change if these constraints do not put a constraint on the multinomial parameter π_1 and so p_D will stay around 1. Only when the constraints on θ_1, θ_2 and θ_3 affect the mobility of the multinomial parameter, p_D (based on π_1) will shrink. On the other hand, p_D based on θ_1, θ_2 and θ_3 will be negative if the constraints were not sufficient to constraint 1. A comparison of the two p_D-values will immediately reveal whether the parameters θ_1, θ_2 and θ_3 are estimable or not.

From a practical point of view we can conclude in general:

1. DIC and p_D should be evaluated in the posterior mean of the multinomial probabilities, and in the posterior mean of the parameters of the model. In WinBUGS language, the latter are called parent nodes. Thus, we need two evaluations of DIC and

p_D, one within WinBUGS and one outside WinBUGS.

2. Only when the two p_D-values are smaller or equal to 2^h-1, there is hope that the prevalence of the disease can be estimated;
3. Models with a high value for DIC indicate a bad model in a Bayesian sense thus where either the model (likelihood) part is badly specified and/or the prior distributions are not compatible with the data. Consequently, when comparing different prior knowledge combined with the same likelihood, prior knowledge that is in conflict with the actually observed data is reflected in a high value for DIC.

5.2 Bayesian P-value

When the model has been overspecified, the Bayesian P-value (as defined in our approach) will be around 0.50. The reason for this is that the posterior probability for the multinomial probabilities will be flat. However this test quantity is a useful indicator for the actual model fit since the Bayesian P-value tends to zero if there is a good model fit and to one if the fit is poor.

The sampled value of δ, $\tilde{\delta}$, is given by

$$\tilde{\delta} = 2\sum_i r_i \ln \frac{r_i}{\hat{r}_i} - 2\sum_i \tilde{r}_i \ln \frac{\tilde{r}_i}{\hat{r}_i} \tag{3}$$

where r_i is the observed cell frequency, $\hat{r}_i$ represents the fitted frequency under the assumed model, i.e. $\hat{\pi}_i N$ (i=1, ...,k) and $\tilde{r}_i$ is the sampled cell frequency under the assumed model. The first part of $\tilde{\delta}$ becomes zero when $\hat{r}_i = r_i$. This happens when constraints on the parent nodes of the π_i are tight in the correct manner ensuring that $\hat{r}_i$ tends to ri. The second part of $\tilde{\delta}$ becomes zero when $\hat{r}_i = \tilde{r}_i$ (i=1, ...,k), which has a probability equal to:

$$\frac{N!}{r_1!\, r_2! \,\cdots\, r_k!}\, \pi_1^{r_1} \pi_2^{r_2} \,\cdots\, \pi_k^{r_k} \tag{4}$$

Assuming $\hat{r}_i = r_i$ (first part of $\tilde{\delta} = 0$), $\tilde{\delta}$ is non-negative when the second part equals 0 and only then there is a contribution to the Bayesian P-value. Thus, the posterior mean $\frac{1}{T}\sum_{t=1}^{T} I\left(\tilde{\delta}_t > 0\right)$ goes to $\frac{N!}{r_1!\, r_2! \,\cdots\, r_k!}\, \pi_1^{r_1} \pi_2^{r_2} \,\cdots\, \pi_k^{r_k}$.

This value is the lower limit attainable by the Bayesian P-value. It is attained when the parent nodes of the multinomial probabilities are sufficiently constrained to assure virtually constant sampling of the multinomial probabilities equal to $\hat{\pi}_i \simeq \frac{r_i}{N}$ ensuring that $\hat{r}_i = r_i$.

In the case of a (very) poorly fitting model, the first part of $\tilde{\delta}$ (= deviance of the observations) (almost) always exceeds the second part (= deviance of sampled values) so that $I(\tilde{\delta}_t > 0)$ (almost) always takes the value 1. Thus the Bayesian P-value tends to 1.

5.3 Modelling exercise

The behaviour of DIC, p_D and the Bayesian P-value will be examined using theoretical frequencies.

The prevalence of the disease is taken equal to 0.5. Further, we assume two diagnostic tests T_1 and T_2, (h = 2), both with specificity equal to 1, i.e. there are no false positive results. The sensitivity of T_1 equals 0.60 and the sensitivity of test T_2 equals 0.70, but there is conditional dependence, i.e. in terms of the parameters in Table 3 θ_4 and θ_5 are not equal. Summarized: $\theta_1 = 0.50$, $\theta_2 = 0.60$, $\theta_3 = 1$, $\theta_4 = 0.90$, $\theta_5 = 0.40$, $\theta_6 = 1$ and $\theta_7 = 1$. This yields the following theoretical probabilities for the 2x2 collapsed contingency table: P(00) = 0.62, P(01) = 0.08, P(10) = 0.03 and P(11) = 0.27. For a study of size 1000 (= N), the expected cell frequencies are therefore $r_1 = 620$, $r_2 = 80$, $r_3 = 30$ and $r_4 = 270$ and the expected number of diseased subjects is equal to $N_{D^+} = 500$. On this data set the following models were tested:

- M1: no prior constraints;
- M2: specificity of $T_1 = 1$, specificity of T_2 =1;
- M3: specificity of $T_1 = 1$, specificity of T_2 =1, sensitivity of T_1 is constrained uniformly to interval [0.5,0.7] and the sensitivity of T_2 is constrained by a uniform prior on θ_4 to interval [0.8,1] and a uniform prior on θ_5 to interval [0.3, 0.5];
- M4: specificity of $T_1 = 1$, specificity of T_2 =1, the sensitivity of T_1 is severely constrained uniformly to interval [0.5999,0.6001] and the sensitivity of T_2 is severely constrained by a uniform

prior on θ_4 to interval [0.8999,0.9001] and a uniform prior on θ_5 to interval [0.3999, 0.4001];

- M5: constraints on specificity and sensitivity of T_1 and T_2 =1 as in M4. Additionally, the prevalence is severely constrained by a uniform prior on θ_1 to interval [0.4999, 0.5001];
- M6: specificity of $T_1 = 1$, specificity of T_2 =1, sensitivity of T_1 is wrongly constrained by a uniform prior to interval [0.8, 1];
- M7: specificity of $T_1 = 1$, specificity of T_2 =1, the sensitivity on T_1 is wrongly constrained by a uniform prior to interval [0.8, 1] and a wrongly positive conditional sensitivity of T_2 by a uniform prior to interval [0.2, 0.4].

The model listings are shown in section A.

In the next section, these models are applied to the 2x2 contingency table of the expected frequencies. This exercise further exemplifies our reasoning in Sections 5.1 and 5.2.

5.4 Results and discussion

The results of applying models M1 to M7 are summarised in Table 6. Note that DIC and p_D calculated from the multinomial probabilities for models M1, M2 and M3 differ only by random MCMC sampling variation.

In models M1 and M2 the constraints are not sufficient to estimate the parameters θ_1 to θ_7 of Table 3. This is reflected by negative p_D-values estimated from the parent nodes. Observe that p_D as calculated from the multinomial probabilities is practically equal to 3, the true value. Further, for both models the Bayesian P-value is about 0.5 indicating no particular problem. Clearly, for both models the prevalence of the disease is overestimated. The constraint imposed on model M3, brings the parent-node p_D close to 3 indicating that now all parameters are estimable. The prevalence is well estimated now and the estimated sensitivities are close to their true values. In models M4 and M5 the constraints are made more stringent but in the correct manner. Model M4 has the lowest DIC value of the two with the lowest p_D-value almost equal to 0 which implies that actually are parameters are set to their correct values. The Bayesian P-value indicates indeed a nearly perfect but non-stochastic model. Further, the prevalence and

the sensitivities are basically equal to their true values. In models M5 and M6 enough constraints have been put on the parameters since for each model the two corresponding p_D-values are almost equal to each other. However, the Bayesian P-values indicate badly fitted models, which is also reflected in a badly estimated prevalence and sensitivities (of course this would not be recognized in practice by the user).

		Parent nodes		Multinomial			Test 1		Test 2	
Model	Bayes-P	DIC	p_D	DIC	p_D	Prev	Se	Sp	Se	Sp
1	0.4916	-90.177	-111.609	24.283	2.936	0.5253	0.3229	1	0.3957	1
2	0.4930	8.303	-13.065	24.342	2.952	0.5688	0.5732	1	0.6680	1
3	0.4793	24.176	2.873	24.279	2.917	0.5058	0.5956	1	0.6939	1
4	0.1852	20.407	0.990	20.471	1.021	0.5006	0.6000	1	0.7000	1
5	0.0004	18.426	0.000	18.445	0.007	0.5000	0.6000	1	0.7000	1
6	0.7000	25.524	2.351	25.468	2.338	0.3850	0.8106	1	0.9049	1
7	1.0000	355.977	1.406	355.899	1.382	0.3848	0.8103	1	0.5006	1

Table 6: Results of the different models using the theoretical data of Section 5.3. For each model the posterior mean of the parameters are given. The column Parent nodes indicate that the calculations were done within WinBUGS and are based on the parameters θ_1 to θ_7 in Table 3. The column Multinomial indicates that the calculations were done outside WinBUGS and are based on the multinomial probabilities.

6 Application of model (2) to field data

6.1 The problem and the data

Porcine cysticercosis is a major problem in many areas, being a debilitating and potentially lethal zoonosis[24,25]. Relatively accurate estimates of prevalence of cysticercae in fattening pigs are essential to appraise the risk for human infection. Several diagnostic tests are used, but none is a gold standard and exact information about test sensitivity and specificity is unavailable. A total of 868 tradition-

ally kept pigs, offered for sale on a market near Lusaka (Zambia) were tested with the following four diagnostic tests: palpation of the tong (TONG), visual inspection of the carcase (VISUAL), an antigen Enzyme-linked Immunosorbent Assay (Ag-ELISA) and an antibody Enzyme-linked Immunosorbent Assay (Ab-ELISA). A summary of the results is shown in Table 7[26].

The data in Table 7 were used to estimate the prevalence and the test characteristics under equation (2) and assuming a variety of expert opinions.

TONG	VISUAL	Ag-ELISA	Ab-ELISA	Number of pigs
0	0	0	0	326
0	0	0	1	42
0	0	1	0	281
0	0	1	1	95
0	1	0	0	0
0	1	0	1	0
0	1	1	0	5
0	1	1	1	4
1	0	0	0	1
1	0	0	1	0
1	0	1	0	2
1	0	1	1	0
1	1	0	0	2
1	1	0	1	1
1	1	1	0	35
1	1	1	1	74

Table 7: Test results of 868 traditional Zambian pigs, subjected to four diagnostic tests (0 = negative test result, 1 = positive test result; TONG = tong palpation; VISUAL = visual carcass inspection; Ag-ELISA = antigen ELISA; Ab-ELISA = antibody ELISA).

6.2 Prior information

'Expert' opinion in the broadest possible sense was used to specify prior information on the diagnostic test characteristics. In this section we will call a model the combination of equation (2) with a particular

set of deterministic and probabilistic (prior information) constraints. Some of the models were constructed from general principles only. For instance, in model M1 the 'expert' opinion states that both test sensitivity and specificity can take any value between zero and one and that the four tests are mutually conditionally independent. For the other models proper expert opinion was used. This expert opinion was obtained from helminthologists at the Institute of Tropical Medicine and at Ghent University. They provided upper and lower limits for the various test sensitivity and specificity values. From biological principles they also concluded that the tests TONG and VISUAL are not independent in a truly infected population, namely a positive test result for TONG is nearly always accompanied by a positive result for VISUAL, i.e. a negative TONG test nearly invariably means a negative VISUAL test. The prior distributions for sensitivity and specificity are taken here as uniform distributions (Beta(1,1)) truncated on the interval [a, b], with a being the under limit and b the upper limit as specified by the experts. These uniform distributions can be replaced by beta-distributions Beta($\alpha\beta$), where α and β are determined such that, say, 95% of the probability mass is located in [a, b].

6.3 Models

Table 8 lists the parameters to be estimated in each of the seven models (M1 to M7) using all four tests that were constructed using the available expert opinion, together with the limits that were applied to each parameter. The starting model (M1) assumes conditional independence of the four tests and no prior information on any of the diagnostic test characteristics (i.e. test sensitivities and specificities have uniform prior distributions on [0, 1]). The model M2 still assumes conditional independence and fixes the specificity of TONG test and the VISUAL test to one, but no other probability constraints were added. The deterministic constraints on model M1 imply that there we are estimating nine parameters when fifteen can be estimated. For model M2 we are estimating seven parameters with again fifteen estimable parameters.

Model M3 again assumes conditional independence, but now probabilistic constraints (inspired by the experts opinions) apply. At face value, there are still seven parameters to be estimated, but the proba-

	M1[a]	M2[b]	M3[c]	M4	M5	M6	M7
θ_1	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1
θ_2	0 - 1	0 - 1	0.3 - 0.7	0 - 1	0 - 1	0 - 1	0 - 1
θ_3	0 - 1	1	1	0 - 1	1	1	1
θ_4	0 - 1	0 - 1	0.8 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1
θ_5	$= \theta_4$	$= \theta_4$	$= \theta_4$	0 - 1	0 - 1	0 - 1	0 - 0.1
θ_6	0 - 1	1	1	0 - 1	1	1	1
θ_7	$= \theta_6$	———	———	0 - 1	———	———	———
θ_8	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1
θ_9	$= \theta_8$	$= \theta_8$	$= \theta_8$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{10}	$= \theta_8$	$= \theta_8$	$= \theta_8$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{11}	$= \theta_8$	$= \theta_8$	$= \theta_8$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{12}	0 - 1	0 - 1	0.95 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1
θ_{13}	$= \theta_{12}$	———	———	0 - 1	———	———	———
θ_{14}	$= \theta_{12}$	———	———	0 - 1	———	———	———
θ_{15}	$= \theta_{12}$	———	———	0 - 1	———	———	———
θ_{16}	0 - 1	0 - 1	0.92 - 1	0 - 1	0 - 1	0 - 1	0 - 1
θ_{17}	$= \theta_{16}$	$= \theta_{16}$	$= \theta_{16}$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{18}	$= \theta_{16}$	$= \theta_{16}$	$= \theta_{16}$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{19}	$= \theta_{16}$	$= \theta_{16}$	$= \theta_{16}$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{20}	$= \theta_{16}$	$= \theta_{16}$	$= \theta_{16}$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{21}	$= \theta_{16}$	$= \theta_{16}$	$= \theta_{16}$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{22}	$= \theta_{16}$	$= \theta_{16}$	$= \theta_{16}$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{23}	$= \theta_{16}$	$= \theta_{16}$	$= \theta_{16}$	0 - 1	0 - 1	0 - 1	0 - 1
θ_{24}	0 - 1	0 - 1	0.98 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1
θ_{25}	$= \theta_{24}$	———	———	0 - 1	0 - 1	0 - 1	0 - 1
θ_{26}	$= \theta_{24}$	———	———	0 - 1	———	———	———
θ_{27}	$= \theta_{24}$	———	———	0 - 1	———	———	———
θ_{28}	$= \theta_{24}$	———	———	0 - 1	———	———	———
θ_{29}	$= \theta_{24}$	———	———	0 - 1	———	———	———
θ_{30}	$= \theta_{24}$	———	———	0 - 1	———	———	———
θ_{31}	$= \theta_{24}$	———	———	0 - 1	———	———	———

[a] equivalent to	TONG	VISUAL	Ag-ELISA	Ab-ELISA
sensitivity	0 - 1	0 - 1	0 - 1	0 - 1
specificity	0 - 1	0 - 1	0 - 1	0 - 1

[b] equivalent to	TONG	VISUAL	Ag-ELISA	Ab-ELISA
sensitivity	0 - 1	0 - 1	0 - 1	0 - 1
specificity	1	1	0 - 1	0 - 1

[c] equivalent to	TONG	VISUAL	Ag-ELISA	Ab-ELISA
sensitivity	0.3 0.7	0.8 - 1	0 - 1	0.92 - 1
specificity	1	1	0.95 - 1	0.98 - 1

Table 8: Parameters to be estimated in the seven models that were constructed from the available expert opinion. $\theta_1 \ldots \theta_{31}$: see Table 3 for the parameter definition; a-b denotes that a is the under- and b is the upper limit of the parameter interval; = [a]: value equal to parameter a in brackets; ———: not to be estimated.

bility constraints imply probabilistic relationships among the parameters and hence fewer parameters need to be estimated. The actual number of parameters estimated in the model should be reflected in the value of p_D. The remaining models all considered conditional dependence. When no constraints are applied 31 parameters need to be estimated while only fifteen parameters are estimable in the collapsed table (see model M4). Putting the TONG specificity and the VISUAL specificity both to one (model M5) reduces the number of parameters to be estimated to nineteen: conditional probabilities θ_3 and θ_6 become one and all parameters, appearing behind $(1-\theta_3)$ and $(1-\theta_6)$ no longer need to be estimated (i.e. θ_7, θ_{13}, θ_{14}, θ_{15}, θ_{26}, θ_{27}, θ_{28}, θ_{29}, θ_{30}, θ_{31}).

The number of parameters to be estimated was further reduced by constraining both θ_{12} and θ_{24} to [0.9 - 1] (model M6), constraints that are moderate by most standards (a specificity equal to 0.90 is considered a low specificity). Finally the conditional probabilities θ_4 and θ_5 were constrained to respectively [0.9 - 1] and [0 - 0.1] (model M7). The constraints applied in models M6 and M7 are of probabilistic nature and hence imply that the actual number of parameters to be estimated lies below nineteen. Model M6 has between seventeen and nineteen parameters to be estimated. Conditional probabilities θ_4 and θ_5, which are constrained in model M7, reflect the expert opinion that the visual carcass inspection result is highly associated with the result of the tongue palpation. If the two tests are made identical (θ_4==1 and θ_5==0), the minimum number of parameters to be estimated becomes six (assuming three independent tests with specificity of one test equal to one) and the actual number of parameters to be estimated lies between six and nineteen. The listing for Model M7 can be found in Appendix A.

6.4 Results

Not all models converged, as expected. Table 9 shows the value of DIC, p_D and the Bayesian P-value for each converged model. Table 10 shows the posterior means together with the 95% credibility intervals of the prevalence and the test characteristics of the four tests.

Model M1 did not converge in WinBUGS, which is not surprising, given that symmetry yields several possible solutions, depending on the starting conditions: replacing sensitivity by the complement of

Model	DIC	p_D	P
M2	97.1	6.52	1.00
M3	925.1	2.89	1.00
M7	70.3	9.86	0.48

Table 9: Deviance information criterion (DIC), effective number of parameters estimated (p_D) and Bayesian P-value (P) for the models that converged.

		TONG		VISUAL		Ag-ELISA		Ab-ELISA	
Model	Prev	Se	Sp	Se	Sp	Se	Sp	Se	Sp
M2	.0144 (0.12-0.17)	0.918 (0.86-0.96)	1.000	0.965 (0.93-0.99)	1.000	0.961 (0.92-0.99)	0.495 (0.46-0.53)	0.635 (0.55-0.72)	0.815 (0.79-0.84)
M3	0.246 (0.22-0.28)	0.540 (0.47-0.61)	1.000	0.803 (0.80-0.81)	1.000	0.973 (0.95-0.99)	0.900 (0.90-0.901)	0.903 (0.90-0.91)	0.952 (0.95-0.96)
M7	0.642 (0.54-0.91)	0.210 (0.14-0.26)	1.000	0.221 (0.15-0.27)	1.000	0.867 (0.62-0.98)	0.947 (0.90-0.997)	0.358 (0.26-0.41)	0.917 (0.85-0.99)

Table 10: Posterior mean for the prevalence and the test characteristics together with the 95% C.I. (in parentheses) for the three models that converged. Prev = prevalence, Se = test sensitivity, Sp = test specificity

specificity, specificity by the complement of sensitivity and prevalence by its own complement yields a symmetric solution (there is thus an inherent problem of identifiability). Indeed, constraining the prevalence to either [0-0.5] or [0.5-1] results in convergence and estimates for all parameters (DIC = 63.3, p_D = 0.3). Model M2 converged and yielded estimates for all parameters. The expert opinion used in model M3 did not improve the model fit, on the contrary DIC increased from 97 to 945 and the Bayesian P-value stayed at one. The Bayesian P-values for models M2 and M3 near one suggest a lack-of-fit indicating that conditional independence test does not hold. Models M4, M5 and M6 did not converge most likely because they were over parameterised implying that the constraints were not strict enough to yield identifiable models. Model M7 converged and yielded the minimum DIC and an acceptable Bayesian P-value of 0.48 (the Bayesian P-value tended to zero when strict constrains were applied).

Table 9 shows the effective number of parameters estimated. For model M7 p_D = 9.86. This illustrates that the 6 constraints (deter-

	TONG	VISUAL	Ag-ELISA	Ab-ELISA
Sensitivity	0.16 (5/31) [0.05-0.34]	0.39 (12/31) [0.22-0.58]	0.65 (20/31) [0.45-0.81]	0.45 (14/31) [0.27-0.64]
Specificity	1.00 (34/34) [0.91-1.00]	1.00 (34/34) [0.91-1.00]	0.91 (31/34) [0.76-0.98]	0.88 (30/34) [0.73-0.97]

Table 11: Test characteristic estimates in a group of 65 pigs, dissected experimentally after slaughter (in brackets number of animals tested positive/number of infected animals in the case of sensitivity, number of animals tested negative/number of disease-free animals in the case of specificity; in square brackets 95% confidence interval).

ministic and probabilistic) on the 20 parameters to estimate have more effect than one might initially think. Indeed, model M7 is based on model (2), which is parameterised in a hierarchical manner with conditional probabilities. Constraints on lower order conditional probabilities must have an effect on higher order conditional probabilities. Taking into account conditional dependence between the various diagnostic tests considerably reduces the estimated test sensitivity of both tong palpation and visual carcass inspection and, most importantly, results in a much higher estimate of the true prevalence (Table 10).

6.5 External model validation

Additional data became available later, allowing external validation of the selected model. Namely, an additional 65 pigs were subjected to the four tests and completely dissected out upon slaughter (= gold standard), permitting the ascertainment of the true infection status and thus allowing estimation of the true prevalence as well as the test characteristics. The true prevalence was estimated as 0.48 (31/65) and the estimates of the test characteristics are shown in Table 11.

Clearly, model M7 (Table 10) resulted in parameter estimates that are reasonably close to those obtained from the experimental dissections (Table 11).

7 Discussion

Analysis of data generated by the application one or more diagnostic tests in a specified population invariably entails over-fitting of the data. The number of parameters that have to be estimated always exceeds the number that can be estimated. This can only be resolved by simplifying the model (deterministic constraints) or through the inclusion of expert opinion (probabilistic constraints). In the latter case only a Bayesian approach can incorporate that information. Observe that the Bayesian approach is slowly becoming accepted by the medical community. Yet, everyday practice is a reflection of the Bayesian philosophy. Indeed, when a test is used within a certain population, it is implicitly assumed that the values of sensitivity and specificity, as supplied by the manufacturer of the test kit, apply to the population studied. Thus, the prior knowledge of the test characteristics is given so much weight that the actual data at hand become insignificant as far as these characteristics are concerned, therefore allowing estimation of the true prevalence.

The model developed on the basis of conditional probabilities allows formalisation of this expert opinion, whatever form it might take. Anything from genuine information acquired through high-quality data to a personal opinion can be quantified and fed as a prior belief probability distribution into the model. Whether it is easy to specify a prior opinion on a conditional probability will depend on the actual tests involved, however we argue that it is practically impossible to give reliable prior information on the sensitivity of a diagnostic test. The user can monitor the effect of this prior belief on the results and it may be easier for the user to appreciate the fact that the actual interpretation of the test results is conditional on the prior opinion. The effect of imposing deterministic and/or probabilistic constraints is reflected in the value of p_D and can thus be evaluated.

Our approach is in sharp contrast to the approach of Pouillot et al.[27] where conditional independence is accepted when a specific test shows no indication against this assumption. However, not much is known about the power of this test. Instead we suggest to work under the assumption of conditional dependence and apply a sensitivity analysis on the estimation of the prevalence and the test characteristics by varying the prior distributions.

The results of the different scenarios applied to the present example

clearly show that the estimate of the infection prevalence depends on the model chosen and that widely varying estimates can be obtained. It is important that users do understand this and realise that the expert opinion has a great impact on the final estimation of the prevalence. However, as the simulation study (and the real-life study) show, DIC, p_D and the Bayesian P-value are useful in the process of selecting a model. We must however warn the user that the information in the collapsed table over the disease groups contains inherently little information on the prevalence and the test characteristics. Finally, the present example shows that classical testing with one or more tests, assuming constancy of test parameters and independence of tests may grossly underestimate true prevalence and thus in our case the seriousness of the zoonosis.

Acknowledgements

The authors wish to thank the anonymous reviewer for her/his constructive comments on an earlier version of this paper.

References

[1] Greiner, M. and Gardner, I.A., (2000). Epidemiological issues in the validation of veterinary diagnostic tests. Preventive Veterinary Medicine 45, 3-22.

[2] Gardner, I.A., Stryhn, H., Lind, P. and Collins, M.T. (2000). Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. Preventive Veterinary Medicine 45, 107-122.

[3] Dendukuri, N. and Joseph, L. (2001). Bayesian approaches to modelling the conditional dependence between multiple diagnostic tests. Biometrics 57, 208-217.

[4] Spiegelhalter, D.J. Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society B 64, 583-640.

[5] Rogan, W.J. and Gladen, B. (1978). Estimating prevalence from the results of a screening test. American Journal of Epidemiology 107, 71-76.

[6] Thrusfield, D., (1995). Veterinary Epidemiology. Oxford University Press, Oxford.

[7] Toma, B., Bnet, J.-J., Dufour, B., Eloit, M., Moutou, F., Sanaa, M. (1991). Glossaire dpidmiologie Animale. ditions du Point Vtrinaire, Maisons-Alfort.

[8] Billiouw, M., Vercruysse, J., Marcotty, T., Speybroeck, N., Chaka, G. and Berkvens, D. (2002). Theileria parva epidemics: a case study in eastern Zambia. Veterinary Parasitology 107, 51-63.

[9] Billiouw, M., Brandt, J., Vercruysse, J., Speybroeck, N., Marcotty, T., Mulumba, M. and Berkvens, D. (2005). Evaluation of the Indirect Fluorescent Antibody Test as a diagnostic tool for East Coast fever in eastern Zambia. Veterinary Parasitology, 127, 189-198.

[10] Saegerman, C., De Waele, L., Gilson, D., Godfroid, J., Thiange, P., Michel, P., Limbourg, D., Vo, T.K.-O., Limet, J., Letesson, J.-J. and Berkvens, D. (2004). Evaluation of three serum i-ELISAs using monoclonal antibodies and protein G as peroxidase conjugate for the diagnosis of bovine brucellosis. Veterinary Microbiology, 100, 91-105.

[11] Begg, C.B (1987). Biases in the assessment of diagnostic tests. Statistics in Medicine, 6, 411–423

[12] Georgiadis, M., Johnson, W., Gardner, I, and Singh, R. (2003). Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. Applied Statistics, 52, 63-76.

[13] Hui, S.L. and Walter, S.D. (1980). Estimating the error rates of diagnostic tests. Biometrics 36, 167-171.

[14] Joseph, L., Gyorkos, T.W. and Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. American Journal of Epidemiology 141, 263-273.

[15] Johnson, W.O., Gastwirth, J.L. and Pearson, L.M. (2000). Screening without a Gold Standard: The Hui-Walter Paradigm Revisited. American Journal of Epidemiology 153, 921-924.

[16] Toft, N., Jrgensen, E. and Hjsgaard, S. (2005). Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Preventive Veterinary Medicine, 68, 19-33.

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

[18] Branscum, A.J., Gardner, I.A. and Johnson, W.O. (2005). Estimation of diagnostic-test sensitivity and specificity through Bayesian modelling. Preventive Veterinary Medicine, 68, 145-163.

[19] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003). Bayesian Data Analysis, Second Editions. Chapman & Hall/CRC, Boca Raton.

[20] Spiegelhalter, D.J., Thomas, A., Best, N.G. and Lunn, D. (2003). WinBUGS Version 1.4 User Manual. MRC Biostatistics Unit, Cambridge.

[21] Adel, A. and Berkvens, D. (2002). Extension of the classic multi-test model, involving covariance between the tests, to a three- and four-test situation. Internal Technical Report, ITM, 16.04.2002.

[22] Young, K and Pettit, L. (1996). Measuring discordancy between prior and data. J. R. Statistical Society, Series B, 58, 679-689.

[23] McCullagh, P. and Nelder, J. (1989). Generalized Linear models. Chapman & Hall, London.

[24] Garcia, H.H. and Del Brutto, O.H. (2000). Taenia solium cysticercosis. Infectious Disease Clinics of North America 14, 97-119.

[25] Phiri, I.K., Dorny, P., Gabriel, S., Willingham, A.L.III, Speybroeck, N. & Vercruysse J. (2002). The prevalence of porcine cysticercosis in Eastern and Southern provinces of Zambia. Veterinary Parasitology 108, 31-39.

[26] Dorny, P., Phiri, I.K., Vercruysse, J., Gabriel, S., Willingham III, A.L., Brandt, J., Victor, B., Speybroeck, N. & Berkvens, D. (2004). A Bayesian approach for estimating values for prevalence and diagnostic test characteristics of porcine cysticercosis. International Journal for Parasitology, 34, 569-576.

[27] Pouillot, R., Gerbier, G. and Gardner, I.A. (2002). TAGS, a program for the evaluation of test accuracy in the absence of a gold standard. Preventive Veterinary Medicine 53, 67-81.

A Appendix A - R and WinBUGS program listings

```
setwd("C:/bayes")
source("diagnostic test function.txt")
results <-c(620,80,30,270)
program <- "diagtest.txt"
doit(results, program)

doit <- function(results, program)
{
## read data
r <- results
N <- sum(r)
data <- list(r=r, N=N)

## let BUGS generate initial values
inits <- NULL

## create list of nodes to be monitored
nodes <- c("bayesp", "p", "pr", "se", "sp")

## call WinBUGS and run program diagtest.txt in working directory
diagtest.sim <- bugs(data, inits, nodes, program, n.chains=3,
n.iter=6000, n.burnin=2000, n.thin=1)

## bugs() calculates the optimistic DIC and pD, we want the
## actual values as calculated by BUGS to assess identifiability
## of the model

## general stuff
dbar <- mean$deviance
numClass <- length(r)

## proportions are calculated from posterior means parent nodes
pr[1] <- mean$p[1]*(1-mean$p[2])*(1-mean$p[5])
        + (1-mean$p[1])*mean$p[3]*mean$p[6]
pr[2] <- mean$p[1]*(1-mean$p[2])*mean$p[5]
        + (1-mean$p[1])*mean$p[3]*(1-mean$p[6])
pr[3] <- mean$p[1]*mean$p[2]*(1-mean$p[4])
        + (1-mean$p[1])*(1-mean$p[3])*mean$p[7]
```

```
pr[4] <- mean$p[1]*mean$p[2]*mean$p[4]
        + (1-mean$p[1])*(1-mean$p[3])*(1-mean$p[7])

dhat <- -2*sum(log(seq(N)))
for (i in 1:numClass)
{
z <- max(r[i],1)
dhat <- dhat+2*sum(log(seq(z))) -2*r[i]*log(pr[i])
}
pd <- dbar-dhat
dic <- dhat + 2*pd
print(c("posterior deviance (using parent nodes)
         = ", formatC(dhat, digits=2, format= "f")),quote=FALSE)
print(c("DIC = ", formatC(dic, digits=2, format= "f")), quote = FALSE)
print(c("pD = ", formatC(pd, digits=2, format= "f")), quote = FALSE)
## using posterior means of actual proportions
pr[1] <- mean$pr[1]
pr[2] <- mean$pr[2]
pr[3] <- mean$pr[3]
pr[4] <- mean$pr[4]

dhat <- -2*sum(log(seq(N)))
for (i in 1:numClass)
{
z <- max(r[i],1)
dhat <- dhat+2*sum(log(seq(z))) -2*r[i]*log(pr[i])
}
pd <- dbar-dhat
dic <- dhat + 2*pd
print(c("posterior deviance (using post. means)
         = ", formatC(dhat, digits=2, format= "f")),quote=FALSE)
print(c("DIC = ", formatC(dic, digits=2, format= "f")), quote = FALSE)
print(c("pD = ", formatC(pd, digits=2, format= "f")), quote = FALSE)

}

model
{
r[1:4] ~ dmulti( p[1:4], N )

p[1] <- theta[1]*(1-theta[2])*(1-theta[5])+(1-theta[1])*theta[3]*theta[6]
```

```
p[2] <- theta[1]*(1-theta[2])*theta[5]+(1-theta[1])*theta[3]*(1-theta[6])
p[3] <- theta[1]*theta[2]*(1-theta[4])+(1-theta[1])*(1-theta[3])*theta[7]
p[4] <- theta[1]*theta[2]*theta[4]+(1-theta[1])*(1-theta[3])*(1-theta[7])

theta[1] ~ dbeta(1,1)
theta[2] ~ dbeta(1,1)
theta[3] ~ dbeta(1,1)
theta[4] ~ dbeta(1,1)
theta[5] ~ dbeta(1,1)
theta[6] ~ dbeta(1,1)
theta[7] ~ dbeta(1,1)

r2[1:4] ~ dmulti( p[1:4], N)
for (i in 1:4)
{
z1[i] <- equals(0,p[i])
y1[i] <- max(z1[i],p[i])
x1[i] <- max(r[i],1)
d[i] <- r[i]*log(x1[i]/(y1[i]*N))

z2[i] <- equals(0,p[i])
y2[i] <- max(z2[i],p[i])
x2[i] <- max(r2[i],1)
d2[i] <- r2[i]*log(x2[i]/(y2[i]*N))
}
G0 <- 2 * sum(d[])
Gt <- 2 * sum(d2[])
bayesp <- step(G0 - Gt)

prev  <- theta[1]
se[1] <- theta[2]
stheta[1] <- theta[3]
se[2] <- theta[2]*theta[4]+(1-theta[2])*theta[5]
stheta[2] <- theta[3]*theta[6]+(1-theta[3])*theta[7]

}
```

Program listings of simulations

```
M1: no prior information at all
Model {
r[1:4] ~ dmulti( p[1:4], N )
p[1] <- theta[1]*(1-theta[2])*(1-theta[5])
         +(1-theta[1])*theta[3]*theta[6]
p[2] <- theta[1]*(1-theta[2])*theta[5]
         +(1-theta[1])*theta[3]*(1-theta[6])
p[3] <- theta[1]*theta[2]*(1-theta[4])
         +(1-theta[1])*(1-theta[3])*theta[7]
p[4] <- theta[1]*theta[2]*theta[4]
         +(1-theta[1])*(1-theta[3])*(1-theta[7])
theta[1] ~ dbeta(1,1)
theta[2] ~ dbeta(1,1)
theta[3] ~ dbeta(1,1)
theta[4] ~ dbeta(1,1)
theta[5] ~ dbeta(1,1)
theta[6] ~ dbeta(1,1)
theta[7] ~ dbeta(1,1)
r2[1:4] ~ dmulti( p[1:4], N)
for (i in 1:4){
z1[i] <- equals(0,p[i])
y1[i] <- max(z1[i],p[i])
x1[i] <- max(r[i],1)
d[i] <- r[i]*log(x1[i]/(y1[i]*N))
z2[i] <- equals(0,p[i])
y2[i] <- max(z2[i],p[i])
x2[i] <- max(r2[i],1)
d2[i] <- r2[i]*log(x2[i]/(y2[i]*N))
}
G0 <- 2 * sum(d[])
Gt <- 2 * sum(d2[])
bayesp <- step(G0 - Gt)
prev   <- theta[1]
se1 <- theta[2]
sp1 <- theta[3]
se2 <- theta[2]*theta[4]+(1-theta[2])*theta[5]
sp2 <- theta[3]*theta[6]+(1-theta[3])*theta[7]
}
```

```
M2: specificity of test1 == 1, specificity of test2 ==1
p[1] <- theta[1]*(1-theta[2])*(1-theta[5])+(1-theta[1])
p[2] <- theta[1]*(1-theta[2])*theta[5]
p[3] <- theta[1]*theta[2]*(1-theta[4])
p[4] <- theta[1]*theta[2]*theta[4]
theta[1] ~ dbeta(1,1)
theta[2] ~ dbeta(1,1)
theta[4] ~ dbeta(1,1)
theta[5] ~ dbeta(1,1)


M3: specificity of test1 == 1,
         specificity of test2 ==1,
         sensitivity test1 constrained,
         sensitivity test2 constrained
p[1] <- theta[1]*(1-theta[2])*(1-theta[5])+(1-theta[1])
p[2] <- theta[1]*(1-theta[2])*theta[5]
p[3] <- theta[1]*theta[2]*(1-theta[4])
p[4] <- theta[1]*theta[2]*theta[4]
theta[1] ~ dbeta(1,1)
theta[2] ~ dbeta(1,1)I(0.5,0.7)
theta[4] ~ dbeta(1,1)I(0.8,1)
theta[5] ~ dbeta(1,1)I(0.3,0.5)


M4: specificity of test1 == 1,
specificity of test2 ==1,
sensitivity test1 severely constrained,
sensitivity test2 severely constrained
p[1] <- theta[1]*(1-theta[2])*(1-theta[5])+(1-theta[1])
p[2] <- theta[1]*(1-theta[2])*theta[5]
p[3] <- theta[1]*theta[2]*(1-theta[4])
p[4] <- theta[1]*theta[2]*theta[4]
theta[1] ~ dbeta(1,1)
theta[2] ~ dbeta(1,1)I(0.5999,0.6001)
theta[4] ~ dbeta(1,1)I(0.8999,0.9001)
theta[5] ~ dbeta(1,1)I(0.3999,0.4001)



M5: specificity of test1 == 1,
```

```
          specificity of test2 ==1,
          sensitivity test1 severely constrained,
          sensitivity test2 severely constrained,
          prevalence severely constrained
p[1] <- theta[1]*(1-theta[2])*(1-theta[5])+(1-theta[1])
p[2] <- theta[1]*(1-theta[2])*theta[5]
p[3] <- theta[1]*theta[2]*(1-theta[4])
p[4] <- theta[1]*theta[2]*theta[4]
theta[1] ~ dbeta(1,1)I(0.4999,0.5001)
theta[2] ~ dbeta(1,1)I(0.5999,0.6001)
theta[4] ~ dbeta(1,1)I(0.8999,0.9001)
theta[5] ~ dbeta(1,1)I(0.3999,0.4001)
```

```
M6: specificity of test1 == 1,
          specificity of test2 ==1,
          sensitivity test1 constrained, but wrong
p[1] <- theta[1]*(1-theta[2])*(1-theta[5])+(1-theta[1])
p[2] <- theta[1]*(1-theta[2])*theta[5]
p[3] <- theta[1]*theta[2]*(1-theta[4])
p[4] <- theta[1]*theta[2]*theta[4]
theta[1] ~ dbeta(1,1)
theta[2] ~ dbeta(1,1)I(0.8,1)
theta[4] ~ dbeta(1,1)
theta[5] ~ dbeta(1,1)
```

```
M7: specificity of test1 == 1,
          specificity of test2 ==1,
          sensitivity test1 constrained, but wrong,
          positive conditional sensitivity test2 constrained, but wrong
p[1] <- theta[1]*(1-theta[2])*(1-theta[5])+(1-theta[1])
p[2] <- theta[1]*(1-theta[2])*theta[5]
p[3] <- theta[1]*theta[2]*(1-theta[4])
p[4] <- theta[1]*theta[2]*theta[4]
theta[1] ~ dbeta(1,1)
theta[2] ~ dbeta(1,1)I(0.8,1)
theta[4] ~ dbeta(1,1)I(0.2,0.4)
theta[5] ~ dbeta(1,1)
```

```
Cysticercosis listing model M7

model
{
r[1:16] ~ dmulti( p[1:16], n )

p[1] <- theta[1]*(1-theta[2])*(1-theta[5])*(1-theta[11])
*(1-theta[23])+(1-theta[1])*theta[3]*theta[6]*theta[12]*theta[24]
p[2] <- theta[1]*(1-theta[2])*(1-theta[5])*(1-theta[11]
)*theta[23]+(1-theta[1])*theta[3]*theta[6]*theta[12]*(1-theta[24])
p[3] <- theta[1]*(1-theta[2])*(1-theta[5])*theta[11]
*(1-theta[22])+(1-theta[1])*theta[3]*theta[6]*(1-theta[12])*theta[25]
p[4] <- theta[1]*(1-theta[2])*(1-theta[5])*theta[11]
*theta[22]+(1-theta[1])*theta[3]*theta[6]*(1-theta[12])*(1-theta[25])
p[5] <- theta[1]*(1-theta[2])*theta[5]*(1-theta[10])*(1-theta[21])
p[6] <- theta[1]*(1-theta[2])*theta[5]*(1-theta[10])*theta[21]
p[7] <- theta[1]*(1-theta[2])*theta[5]*theta[10]*(1-theta[20])
p[8] <- theta[1]*(1-theta[2])*theta[5]*theta[10]*theta[20]
p[9] <- theta[1]*theta[2]*(1-theta[4])*(1-theta[9])*(1-theta[19])
p[10]<- theta[1]*theta[2]*(1-theta[4])*(1-theta[9])*theta[19]
p[11]<- theta[1]*theta[2]*(1-theta[4])*theta[9]*(1-theta[18])
p[12]<- theta[1]*theta[2]*(1-theta[4])*theta[9]*theta[18]
p[13]<- theta[1]*theta[2]*theta[4]*(1-theta[8])*(1-theta[17])
p[14]<- theta[1]*theta[2]*theta[4]*(1-theta[8])*theta[17]
p[15]<- theta[1]*theta[2]*theta[4]*theta[8]*(1-theta[16])
p[16]<- theta[1]*theta[2]*theta[4]*theta[8]*theta[16]

for (i in 1:16)
{
d[i] <- r[i]*log(max(r[i],1)/(p[i]*n))
}
G0 <- 2 * sum(d[])

r2[1:16] ~ dmulti(p[1:16], n)

for (i in 1:16)
{
```

```
d2[i] <- r2[i]*log(max(r2[i],1)/(p[i]*n))
}
Gt <- 2 * sum(d2[])

bayesp <- step(G0 - Gt)

theta[1]  ~ dbeta(1,1)
theta[2]  ~ dbeta(1,1)
theta[3]  <- 1
theta[4]  ~ dbeta(1,1)I(.9,1)
theta[5]  ~ dbeta(1,1)I(0,.1)
theta[6]  <- 1
theta[8]  ~ dbeta(1,1)
theta[9]  ~ dbeta(1,1)
theta[10] ~ dbeta(1,1)
theta[11] ~ dbeta(1,1)
theta[12] ~ dbeta(1,1)I(.9,1)
theta[16] ~ dbeta(1,1)
theta[17] ~ dbeta(1,1)
theta[18] ~ dbeta(1,1)
theta[19] ~ dbeta(1,1)
theta[20] ~ dbeta(1,1)
theta[21] ~ dbeta(1,1)
theta[22] ~ dbeta(1,1)
theta[23] ~ dbeta(1,1)
theta[24] ~ dbeta(1,1)I(.9,1)
theta[25] ~ dbeta(1,1)
}
list(r=c(326,42,281,95,0,0,5,4,1,0,2,0,2,1,35,74), n = 868)
```

B Appendix B - Modelling conditional dependence between multiple diagnostic tests, using co-variances between test results

Table 12 presents the probability model, as developed by Gardner at al.[2] and Dendukuri & Joseph[3], used to calculate the probabilities in function of test sensitivities and specificities, when two diagnostic tests are applied to all subjects of a population.

Probability	Expected value
	Infected animals
$\Pr(T_1^- \cap T_2^-)$	$(1-Se_1)(1-Se_2) + cov_{Se_1Se_2}$
$\Pr(T_1^- \cap T_2^+)$	$(1-Se_1)Se_2 - cov_{Se_1Se_2}$
$\Pr(T_1^+ \cap T_2^-)$	$Se_1(1-Se_2) - cov_{Se_1Se_2}$
$\Pr(T_1^+ \cap T_2^+)$	$Se_1Se_2 + cov_{Se_1Se_2}$
	Disease-free animals
$\Pr(T_1^- \cap T_2^-)$	$Sp_1Sp_2 + cov_{Sp_1Sp_2}$
$\Pr(T_1^- \cap T_2^+)$	$Sp_1(1-Sp_2) - cov_{Sp_1Sp_2}$
$\Pr(T_1^+ \cap T_2^-)$	$(1-Sp_1)Sp_2 - cov_{Sp_1Sp_2}$
$\Pr(T_1^+ \cap T_2^+)$	$(1-Sp_1)(1-Sp_2) + cov_{Sp_1Sp_2}$
	Total population
$\Pr(T_1^- \cap T_2^-)$	$\Pr(D^+)[(1-Se_1)(1-Se_2) + cov_{Se_1Se_2}]$ $+[1-\Pr(D^+)][Sp_1Sp_2 + cov_{Sp_1Sp_2}]$
$\Pr(T_1^- \cap T_2^+)$	$\Pr(D^+)[(1-Se_1)Se_2 - cov_{Se_1Se_2}]$ $+[1-\Pr(D^+)][Sp_1(1-Sp_2) - cov_{Sp_1Sp_2}]$
$\Pr(T_1^+ \cap T_2^-)$	$\Pr(D^+)[Se_1(1-Se_2) - cov_{Se_1Se_2}]$ $+[1-\Pr(D^+)][(1-Sp_1)Sp_2 - cov_{Sp_1Sp_2}]$
$\Pr(T_1^+ \cap T_2^+)$	$\Pr(D^+)[Se_1Se_2 + cov_{Se_1Se_2}]$ $+[1-\Pr(D^+)][(1-Sp_1)(1-Sp_2) + cov_{Sp_1Sp_2}]$

Table 12: Observed test results, when applying two tests to each individual, in function of test characteristics. $T_i^{\cdots}$: result of test i; Se_i, Sp_i: sensitivity, specificity of test i; $cov_{Se_1Se_2}$: test result co variance within diseased animal population; $cov_{Sp_1Sp_2}$: test result co variance within disease-free population; $\Pr(D^+)$: true prevalence.

The above results can be expanded to three or more tests.

$$cov_{Se_1Se_2Se_3} = \frac{1}{n}\sum_{i=1}^{n}(T_1^i - \bar{T_1})(T_2^i - \bar{T_2})(T_3^i - \bar{T_3}) \tag{5}$$

In the case of $\Pr(T_1^+T_2^+T_3^+|D^+)$, we are interested in animals that test positive. Thus, T_j^i are replaced by 1 in the case of positive test result, by 0 in the case of a negative test result. The averages are in this case the test sensitivities, the probability to test positive or the average number of positive results. Thus, $\bar{T}_j$ are replaced by Se_j, yielding:

$$cov_{Se_1Se_2Se_3} = \frac{1}{n}\sum_{i=1}^{n}(T_1^i - Se_1)(T_2^i - Se_2)(T_3^i - Se_3) \qquad (6)$$

$$\begin{aligned} cov_{Se_1Se_2Se_2} = \tfrac{1}{n}\sum_{i=1}^{n} (&T_1^iT_2^iT_3^i - T_1^iT_2^iSe_3 - T_1^iT_3^iSe_2 - T_2^iT_3^iSe_1 \\ &+T_1^iSe_2Se_3 + T_2^iSe_1Se_2 + T_3^iSe_1Se_2 - Se_1Se_2Se_3) \end{aligned}$$

Knowing that: $\frac{1}{n}\sum_{i=1}^{n}(T_1^i \cdot T_2^i \cdot T_3^i) = \Pr(T_1^+ \cap T_2^+ \cap T_3^+|D^+)$, $\frac{1}{n}\sum_{i=1}^{n}(T_k^iT_l^iSe_m) = \Pr\left(T_k^i \cap T_l^i|D^+\right)Se_m$, $\frac{1}{n}\sum_{i=1}^{n}(T_k^iSe_lSe_m) = \Pr(T_k^+|D^+)Se_lSe_m = Se_kSe_lSe_m = Se_1Se_2Se_3$ and $\frac{1}{n}\sum_{i=1}^{n}(Se_1Se_2Se_3) = Se_1Se_2Se_3$, this becomes:

$$\begin{aligned} cov_{Se_1Se_2Se_3} = \Pr(T_1^+ \cap T_2^+ \cap T_3^+|D^+) - \Pr(T_1^+ \cap T_2^+|D^+)Se_3 - \Pr(T_1^+ \cap T_3^+|D^+)Se_2 \\ - \Pr(T_2^+ \cap T_3^+|D^+)Se_1 + 2Se_1Se_2Se_3 \end{aligned}$$

Replacing $\Pr(T_k^+ \cap T_l^+|D^+)$ by $[Se_kSe_l + cov_{Se_kSe_l}]$ yields:

$$cov_{Se_1Se_2Se_3} = \Pr(T_1^+ \cap T_2^+ \cap T_3^+|D^+) - cov_{Se_1Se_2}Se_3 - cov_{Se_1Se_3}Se_2 - cov_{Se_2Se_3}Se_1 - Se_1Se_2Se_3$$

and finally

$$\Pr(T_1^+ \cap T_2^+ \cap T_3^+|D^+) = Se_1Se_2Se_3 + Se_1cov_{Se_2Se_3} + Se_2cov_{Se_1Se_3} + Se_3cov_{Se_1Se_2} + cov_{Se_1Se_2Se_3} \qquad (7)$$

Other terms are calculated in analogy. If the probability to obtain a negative result is required, T_j^i is replaced by 1 in the case of negative test result, by 0 in the case of a positive test result. The average in this case becomes $(1-Se_j)$. Given that $cov_{Se_kSe_l(1-Se_m)} = -cov_{Se_1Se_2Se_3}$ the following eight probabilities (adding to 1) for infected animals are obtained:

$$\Pr(T_1^+ \cap T_2^+ \cap T_3^+|D^+t) = Se_1Se_2Se_3 + Se_1cov_{Se_2Se_3} + Se_2cov_{Se_1Se_3} + Se_3cov_{Se_1Se_2} + cov_{Se_1Se_2Se_3}$$

$$\Pr(T_1^+ \cap T_2^+ \cap T_3^- |D^+) = Se_1Se_2(1-Se_3) - Se_1cov_{Se_2Se_3} - Se_2cov_{Se_1Se_3} + (1-Se_3)cov_{Se_1Se_2} - cov_{Se_1Se_2Se_3}$$

$$\Pr(T_1^+ \cap T_2^- \cap T_3^+|D^+) = Se_1(1-Se_2)Se_3 - Se_1cov_{Se_2Se_3} + (1-Se_2)cov_{Se_1Se_3} - Se_3cov_{Se_1Se_2} - cov_{Se_1Se_2Se_3}$$

$$\Pr(T_1^+ \cap T_2^- \cap T_3^- | D^+) = Se_1(1-Se_2)(1-Se_3)+Se_1 cov_{Se_2Se_3}-(1-Se_2)cov_{Se_1Se_3}-(1-Se_3)cov_{Se_1Se_2}+cov_{Se_1Se_2Se_3}$$

$$\Pr(T_1^- \cap T_2^+ \cap T_3^+ | D^+) = (1-Se_1)Se_2Se_3+(1-Se_1)cov_{Se_2Se_3}-Se_2cov_{Se_1Se_3}-Se_3cov_{Se_1Se_2}-cov_{Se_1Se_2Se_3}$$

$$\Pr(T_1^- \cap T_2^+ \cap T_3^- | D^+) = (1-Se_1)Se_2(1-Se_3)-(1-Se_1)cov_{Se_2Se_3}+Se_2cov_{Se_1Se_3}-(1-Se_3)cov_{Se_1Se_2}+cov_{Se_1Se_2Se_3}$$

$$\Pr(T_1^- \cap T_2^- \cap T_3^+ | D^+) = (1-Se_1)(1-Se_2)Se_3-(1-Se_1)cov_{Se_2Se_3}-(1-Se_2)cov_{Se_1Se_3}+Se_3cov_{Se_1Se_2}+cov_{Se_1Se_2Se_3}$$

$$\Pr(T_1^- \cap T_2^- \cap T_3^- | D^+) = (1-Se_1)(1-Se_2)(1-Se_3)+(1-Se_1)cov_{Se_2Se_3}+(1-Se_2)cov_{Se_1Se_3}+(1-Se_3)cov_{Se_1Se_2}-cov_{Se_1Se_2Se_3}$$

Equations for disease-free animals are calculated in a similar way: when T_j^i represents a negative result, it is replaced by 1 for a negative result and 0 for a positive result and the average is the test specificity, when it represents a positive result, it is replaced by 1 for a positive result and 0 for a negative result and the average becomes $(1 - Sp_j)$. The probability to test negatively in all three test, given a true disease-free status is:

$$\Pr(T_1^- \cap T_2^- \cap T_3^- | D^-) = Sp_1Sp_2Sp_3+Sp_1 cov_{Sp_2Sp_3}+Sp_2 cov_{Sp_1Sp_3}+Sp_3 cov_{Sp_1Sp_2}+cov_{Sp_1Sp_2Sp_3}$$

The rest of the equations are obtained and both the infected animal part and disease-free part are combined as for the two-test situation by multiplying them respectively by p and $(1 - p)$.

All of this yields the following generalised formula for h diagnostic tests applied to each subject:

$$p_{i_1 i_2 n i_h} = \Pr\left(D^+\right)\cdot\left[\sum_{J|\#J\neq 1}\sigma_J\cdot\left(\prod_t \varphi_J(t)\right)\right]+\left[1-\Pr\left(D^+\right)\right]\cdot\left[\sum_{J|\#J\neq 1}\sigma'_J\cdot\left(\prod_t \xi_J(t)\right)\right]$$

where $T = \{1, 2, \ldots, h\}$; $\forall t \in T \quad (i_t \in \{0, 1\})$; $J \in P(T)$; $\sigma_{12\ldots} = cov_{Se_1Se_2\ldots}$; $\sigma'_{12\ldots} = cov_{Sp_1Sp_2\ldots}$; $\sigma_\emptyset = 1$; $\sigma'_\emptyset = 1$

$$t \quad \mapsto \quad \varphi_J(t) = \begin{cases} -(-1)^{i_t} & \forall t \in J \\ i_t Se_t + (1-i_t)(1-Se_t) & \forall t \notin J \end{cases}$$

$$t \quad \mapsto \quad \xi_J(t) = \begin{cases} (-1)^{i_t} & \forall t \in J \\ (1-i_t)Sp_t + i_t(1-Sp_t) & \forall t \notin J \end{cases}$$

The co-variances are constrained within well-defined limits to ensure that $p_{i_1 i_2 \ldots i_h}$ is confined to the domain (0,1). The limits depend on the values of the sensitivities or specificities. Gardner et al.[2] and Dendukuri & Joseph[3] give the following limits for the two-test case:

$$\begin{array}{rcccl} \max\left[-(1-Se_1)(1-Se_2), -Se_1Se_2\right] & \leq & \sigma_{12} & \leq & \min\left[Se_1(1-Se_2), (1-Se_1)Se_2\right] \\ \max\left[-(1-Sp_1)(1-Sp_2), -Sp_1Sp_2\right] & \leq & \sigma'_{12} & \leq & \min\left[Sp_1(1-Sp_2), (1-Sp_1)Sp_2\right] \end{array} \quad (8)$$

The limits for the three-way co-variance in the case of three tests become:

$$\max\left\{\begin{array}{c} -\left[Se_1Se_2Se_3+Se_1\sigma_{23}+Se_2\sigma_{13}+Se_3\sigma_{12}\right] \\ -\left[\left(1-Se_1\right)\left(1-Se_2\right)Se_3-\left(1-Se_1\right)\sigma_{23}-\left(1-Se_2\right)\sigma_{13}+Se_3\sigma_{12}\right] \\ -\left[\left(1-Se_1\right)Se_2\left(1-Se_3\right)-\left(1-Se_1\right)\sigma_{23}+Se_2\sigma_{13}-\left(1-Se_3\right)\sigma_{12}\right] \\ -\left[Se_1\left(1-Se_2\right)\left(1-Se_3\right)+Se_1\sigma_{23}-\left(1-Se_2\right)\sigma_{13}-\left(1-Se_3\right)\sigma_{12}\right] \end{array}\right\}$$

$$\leq \quad \sigma_{123} \quad \leq$$

$$\min\left\{\begin{array}{c} \left[\left(1-Se_1\right)Se_2Se_3+\left(1-Se_1\right)\sigma_{23}-Se_2\sigma_{13}-Se_3\sigma_{12}\right] \\ \left[Se_1\left(1-Se_2\right)Se_3-Se_1\sigma_{23}+\left(1-Se_2\right)\sigma_{13}-Se_3\sigma_{12}\right] \\ \left[Se_1Se_2\left(1-Se_3\right)-Se_1\sigma_{23}-Se_2\sigma_{13}+\left(1-Se_3\right)\sigma_{12}\right] \\ \left[\left(1-Se_1\right)\left(1-Se_2\right)\left(1-Se_3\right)+\left(1-Se_1\right)\sigma_{23}+\left(1-Se_2\right)\sigma_{13}+\left(1-Se_3\right)\sigma_{12}\right] \end{array}\right\}$$

Limits for the co-variance in the disease-free group are calculated in a similar way (by replacing Se by Sp and σ by σ'. The following inequality represents the generalised h-way co-variance:

$$\max\left\{\left(\sum i_t \bmod 2 \equiv 0\right)\sum_{J=\emptyset|\#J\neq 1}^{h-1}\sigma_J\left[\prod\varphi_J\left(t\right)\right]\right\} \leq \sigma_{12nh} \leq \min\left\{\left(\sum i_t \bmod 2 > 0\right)\sum_{J=\emptyset|\#J\neq 1}^{h-1}\sigma_J\left[\prod\varphi_J\left(t\right)\right]\right\}$$

$$\max\left\{\left(\sum i_t \bmod 2 \equiv 0\right)\sum_{J=\emptyset|\#J\neq 1}^{h-1}\sigma'_J\left[\prod\xi_J\left(t\right)\right]\right\} \leq \sigma'_{12nh} \leq \min\left\{\left(\sum i_t \bmod 2 > 0\right)\sum_{J=\emptyset|\#J\neq 1}^{h-1}\sigma'_J\left[\prod\xi_J\left(t\right)\right]\right\}$$

A first problem with this approach lies in the specification of the initial values for a Bayesian analysis. Indeed the restrictions on the parameters complicate the choice of valid starting values to ensure initial probabilities $p_{i_1 i_2 \ldots i_h}$ between 0 and 1. Secondly, and more importantly, probability distributions for the covariances (i.e. generalised beta distributions) must be defined beforehand, based on expert opinion. Providing this information directly is very difficult. The only sensible way to acquire this information is by phrasing the questions in terms of conditional probabilities and transforming the estimates thus obtained in terms of the co-variance terms.