University Library

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Clarke, SJ;Jones, SA

Title:
Bayesian Estimation for Diagnostic Testing of Biosecurity Risk Material in the Absence of a Gold Standard when Test Data are Incomplete

Date:
2015-09-23

Citation:
Clarke, S. J. & Jones, S. A. (2015). Bayesian Estimation for Diagnostic Testing of Biosecurity Risk Material in the Absence of a Gold Standard when Test Data are Incomplete. Journal of Agricultural Biological and Environmental Statistics, 20 (3), pp.389-408. https://doi.org/10.1007/s13253-015-0214-5.

Persistent Link:
https://hdl.handle.net/11343/283242

Noname manuscript No.
(will be inserted by the editor)

Bayesian estimation for diagnostic testing of biosecurity risk material in the absence of a gold standard when test data are incomplete

Sandy J. Clarke (sjclarke@unimelb.edu.au)

Statistical Consulting Centre, Department of Mathematics and Statistics,

University of Melbourne, Parkville, 3010, Australia

Stuart A. Jones (s.jones@unimelb.edu.au)

Centre of Excellence for Biosecurity Risk Analysis, School of Botany,

University of Melbourne, Parkville, 3010, Australia.

the date of receipt and acceptance should be inserted later

Address(es) of author(s) should be given

Noname manuscript No.
(will be inserted by the editor)

Bayesian estimation for diagnostic testing of biosecurity risk material in the absence of a gold standard when test data are incomplete

the date of receipt and acceptance should be inserted later

Address(es) of author(s) should be given

Abstract

Diagnostic testing is used by biosecurity officers for the detection and identification of plant and animal pathogens, often informing high-consequence decisions such as restricting the entry of trade goods. It is rare that such tests can be considered gold standards; however, uncertainty can be reduced by using the results of other tests, measuring performance on samples of known status and incorporating prior knowledge from expert judgement. This article presents an extension to the methods of Joseph, Gyorkos, and Coupal (1995), and Dendukuri and Joseph (2001) for Bayesian estimation in the absence of a gold standard test, which allows for the use of incomplete test data. This extension is demonstrated with a novel application: the case study of myrtle rust from Holliday et al. (2013), which involves samples from potential biosecurity risk material on importation pathways to Australia. The samples were tested at two laboratories and prior estimates for pathway prevalence were obtained by expert elicitation. The Bayesian estimation was based on a model with and without covariances for the test results to assess the assumption of conditional independence. The results show that pathogen prevalence, diagnostic sensitivity and diagnostic specificity can be estimated using all available data even where some samples have been subject to only one of two available tests. The results also indicate the importance of consideration of the assumption of conditional independence. The findings enable diagnostic testing laboratories and decision makers to make use of all test results and to explicitly incorporate prior knowledge to estimate pathogen prevalence and test accuracy.

KEYWORDS: Biosecurity; disease prevalence; plant pathogen; *Puccinia psidii*; sensitivity; specificity.

1 Introduction

1.1 Diagnostic testing

Diagnostic testing is used routinely by medical doctors, epidemiologists, veterinarians and quarantine officers to infer the presence and absence of diseases and their causal pathogens. Test results may inform the diagnosis of an individual or contribute to the estimation of disease prevalence in a population. These results often support high-consequence decisions such as whether to undertake a medical treatment or restrict entry of a traded commodity.

Despite the importance of diagnostic test results, error-free or gold standard tests are rare and tests with higher accuracy are typically more expensive (Garber and Solomon, 1999; Vijan et al., 2001). As such, it is necessary to be able to make use of the information provided by a fallible test. Estimation can be enhanced by using results from additional (preferably independent) fallible tests, experimental data about test accuracy (sensitivity and specificity) and prior knowledge about the estimated parameters (Dunson, 2001). It is also important that uncertainty is reported for diagnostic test results so decision makers can apply their appetite for risk (Aven, 2013), depending on the relative consequences of false positives and false negatives.

1.2 Testing for plant biosecurity

While diagnostic testing finds its origins in medical practice, the methods also apply in other contexts. The motivating example for our approach is estimation of the presence and prevalence of plant pathogens in the context of biosecurity. Other Bayesian approaches have been applied to cases such as in-situ crop and soil-borne disease (Mila et al., 2003; Makowski et al., 2008; Vidal et al., 2012; Krull et al., 2013), but the approach advocated here has not, nor has it been applied to the testing of biosecurity risk material on import.

Specifically, the case study involves the molecular diagnosis of *Puccinia psidii* sensu lato, which is the causal pathogen of the disease known variously as Euca-

lyptus rust, guava rust, ohia rust and myrtle rust (Coutinho et al., 1998; Mireku and Simpson, 2002; Kawanishi et al., 2009; Carnegie et al., 2010). *P. psidii* s.l. is a fungal pathogen with a wide host range in the Myrtaceae (Simpson, Thomas, and Grgurinovic, 2006) and is a threat to natural plant communities and commercial timber industries (Tommerup, Alfenas, and Old, 2003; Grgurinovic, Walsh, and Macbeth, 2006; Glen et al., 2007). Potential pathways by which the pathogen might enter Australia were identified by Grgurinovic et al. (2006).

For plant pathogens, molecular methods of identification – such as polymerase chain reaction (PCR) assays – are commonly used to verify the identity of a pathogen that has a preliminary identification (a strong prior) based on knowledge of the host plant and disease symptoms. PCR testing can also be used for the detection of a pathogen in 'environmental samples' that have no direct evidence of pathogen presence. Environmental samples typically have indirect evidence of potential pathogen presence (a weak prior) based on predictive modelling or expert judgement.

Positive test results of environmental samples can indicate the presence of a pathogen on an import exposure pathway. The pathways targeted in this case study were cut flowers from Colombia and Ecuador; air passengers from central and south America, Hawaii and Japan; and cargo containers from central and south America. The sampling program used a vacuum to extract particulate material from the surfaces of sampled pathway objects onto 5 μm filter papers (Holliday et al., 2013). Each filter paper was cut into pieces (here termed 'pathway samples') for testing at two plant diagnostic laboratories in Australia (here termed 'Lab A' and 'Lab B') for the presence or absence of *P. psidii* using a PCR test method adapted from Langrell, Glen, and Alfenas (2008).

1.3 The problem of interest

The difficulty with the analysis of the samples in this case study is that the tests applied at Lab A and Lab B demonstrated limited reliability and could not be considered to be diagnostic gold standards. Furthermore, results were not available for the tests from both

labs in every case. Where the results of two tests are used it is common that some data will be missing for either or both of the tests (e.g., Vlems et al., 2003; Schiffman et al., 2005; Vettraino et al., 2010). This is particularly true of environmental samples, which may have been collected for exploratory data collection or monitoring purposes and may not have been subject to both tests due to time or cost constraints.

In this study, the limited reliability of the tests is addressed by blind, randomised testing of process control samples of known infestation status (positive and negative) to estimate the diagnostic sensitivity and specificity of the tests for the pathway samples. These data provide priors for sensitivity and specificity while prior estimates of pathway prevalence are derived from expert judgement. Anecdotal evidence suggests that the use of such protocols and multiple laboratories is not routine in testing environmental samples for plant pathogens, and that sensitivity and specificity are often estimated from relatively idealised conditions, such as testing of target species DNA or testing of samples for which there is a strong prior. This study is consistent with the common case that test accuracy is lower for environmental samples, suggesting the need for appropriate statistical treatment of the results using estimates of diagnostic sensitivity and specificity from process controls. The effectiveness of this approach is increased where all available test results can be used even where some paired sample results are missing.

Depending on the method of statistical analysis, missing data may be handled in a number of ways, such as by excluding the paired data (Vlems et al., 2003; Vettraino et al., 2010) or by conducting a sensitivity analysis for possible values of the missing data (Schiffman et al., 2005). This can result in a reduction in statistical power. Hence, the ability to use all available test results even where paired samples are missing is of value for many practical applications. We seek to demonstrate this in a Bayesian framework using the case study of diagnostic testing for *P. psidii* s.l. in environmental samples. The absence of test results from one lab or the other for some pathway samples is addressed by developing an extension to the statistical analysis method of Joseph et al. (1995). As this method assumes conditional independence between the test results for

the two labs, we applied a further extension of the methods of Dendukuri and Joseph (2001). This article aims to demonstrate Bayesian estimation of disease prevalence with a fallible test, where no gold standard test is available and test data are incomplete.

2 Methods

2.1 Overview

The lack of a gold standard introduces uncertainty in the test results, and necessitates estimation that incorporates this uncertainty. For problems of this nature the approach we propose is a Bayesian one, which combines the available information in one estimation process. This approach requires prior estimates to be made for the prevalence of the disease on each pathway, and for the sensitivity and specificity of the test from each lab. It combines this knowledge, in the form of prior distributions, with the pathway sample data to give a combined posterior distribution for the parameters. Methods of this form have been used in a variety of contexts (Demissie et al., 1998; Enoe, Georgiadis, and Johnson, 2000; Dorny et al., 2004; Speirs-Bridge et al., 2010).

The prior distributions for the prevalence of *P. psidii* s.l. on each pathway were obtained by eliciting expert judgements. The prior distributions for the diagnostic sensitivity and specificity of each lab test were obtained by randomised inclusion of samples with known pathogen status presented as pathway samples (i.e., blind testing of process controls). The sample data are the test results of the pathway samples. Each pathway – flowers, passengers and containers – was modelled separately.

Posterior distributions for each of the parameters are difficult to compute analytically, but marginal posterior distributions can be obtained using Gibbs Sampling (see, for example, Casella and George 1992, for a comprehensive treatment). The main advantage of this approach is that the construction of the prior incorporates the relevant information contained in the test results and enables simultaneous inferences about all parameters, including prevalence and test accuracy. These results were generated us-

ing both the models with and without correlated tests, in order to assess the assumption of conditional independence of tests and determine the best model for interpretation. The more complex model with correlated tests resulted in some marginal posterior distributions of unknown form, so sampling-importance resampling (Rubin, 1988) was used within the Gibbs sampler.

2.2 Definitions

The notation used follows that of Joseph et al. (1995), and Dendukuri and Joseph (2001) with an extension to incorporate samples for which results are only available from one of the two tests. This is the approach used for each pathway.

Table 1 gives the structure of the data. In our example, Tests 1 and 2 refer to the results from Lab A and Lab B, respectively.

[Table 1 about here.]

Let the unobserved latent data $Y_1, \ldots, Y_8$ represent the number of true positive samples out of the observed values $n_1, \ldots, n_8$, respectively, in Table 1.

Table 2 extends Table 4 of Joseph et al. (1995) and details the likelihood contributions of the observed data (n_i) and latent data (Y_i), depending on the true status of the sample. The prevalence is given by π, while S_1, C_1, S_2 and C_2 are the sensitivities and specificities for the two tests, respectively, and *covs* and *covc* are the covariances between the two tests for the positive and negative samples, respectively.

[Table 2 about here.]

2.3 Prior distributions

For independent samples, the prevalence, sensitivity and specificity are all binomial proportions, for which the conjugate prior distribution is a beta distribution.

$$\pi \sim \text{Beta}(\alpha_\pi, \beta_\pi)$$

$$S_1 \sim \text{Beta}(\alpha_{S_1}, \beta_{S_1})$$

$$C_1 \sim \text{Beta}(\alpha_{C_1}, \beta_{C_1})$$

$$S_2 \sim \text{Beta}(\alpha_{S_2}, \beta_{S_2})$$

$$C_2 \sim \text{Beta}(\alpha_{C_2}, \beta_{C_2})$$

For the model with correlated tests, we are only interested in situations where the two tests are positively correlated and the sensitivities themselves provide an upper bound on the covariance, so uniform prior distributions were used for *covs* and *covc* with the following constraints:

$$0 \leq covs \leq \min(S_1, CS_2) - S_1S_2$$

$$0 \leq covc \leq \min(C_1, C_2) - C_1C_2.$$

The prior distribution for π was obtained by eliciting the judgements of informed and expert participants. The participants were asked to provide prior estimates for the proportion of objects carrying *P. psidii* s.l. spores on each of the pathways; i.e., the prevalence of infestation among objects on the pathway. The elicitation was carried out using the four-point elicitation method described in Speirs-Bridge et al. (2010).

There were 13 participants, 7 of whom had expertise in *P. psidii* s.l. through publication or involvement in research on the topic, and 6 of whom had no direct expertise in the pathogen but had experience in research related to biosecurity or conservation. The six additional participants were included to increase the diversity of expert judgement, which is important for improving group performance even where this introduces participants that have a social expectation of lower individual performance (Burgman et al., 2011).

All participants were given general information about the pathogen, including its host range and its geographical range. Information was also provided about the proportion of samples that were collected from each country for that pathway. For

flowers, information was also given about the proportion of each flower type that was sampled, and for passengers, information was given about the proportion of people who had done activities involving contact with vegetation. The information presented to the participants was drawn from the data that were available shortly before the first PCR results were available; the majority of the samples available at this point in time subsequently underwent PCR testing. Thus, judgments of the participants were informed by profile information similar, although not identical, to that of the tested populations.

The results of the elicited lower and upper estimates were scaled linearly to standardised 80% intervals based on the participants' reported confidence, while the best estimates were not adjusted. The pooled group results for each of the lower, upper and best estimates were calculated as simple means, which is a suitable method to buffer against variations in individual performance (McBride, Fidler, and Burgman, 2012). The group results were then converted to α_π and β_π using the methods outlined in Keefer and Bodily (1983).

The prior distribution for the sensitivity and specificity for each lab was obtained using the results of samples with known pathogen status. The sensitivity and specificity can be viewed as estimates of p from a binomial distribution so the number of false and true negatives and positives from each lab could be directly converted into the relevant α and β shape parameters, with α as the number of successes and $(\alpha + \beta)$ as the total sample size.

Uniform genbeta(1,1) prior distributions were used for *covs* and *covc* with the constraints discussed above.

2.4 Gibbs sampler

The conditional distributions for the Gibbs sampler are included in the supplementary materials. The Gibbs sampler alternates between (1) derivation of the posterior distributions for the latent data conditioning on the current prevalence and diagnostic test values and (2) derivation of the prevalence and diagnostic test values conditioning on

the current estimate of the latent data. The estimation of $Y_5, \ldots, Y_8$ will add additional steps to the Gibbs sampler of Joseph et al. (1995), also relying on the estimates of S_1, C_1, S_2 and C_2 at each stage, along with *covs* and *covc* in the dependence case.

After a suitable initialisation period of 500 runs, a total of 20,000 cycles were used to generate the posterior distributions. Convergence of the algorithm appeared to have occurred after 500 runs, as evidenced by selected subsampling of the posterior samples and varying of the number of cycles.

3 Results

3.1 Priors for pathway prevalence

Figure 1 shows elicited expert estimates of the proportion of objects carrying *P. psidii* s.l. spores for each of the three sampled pathways. There is a large amount of uncertainty in the estimates, although the best estimates are mostly clustered around values less than 10%. Thus, the average best estimate for all three pathways is around 10%, with upper bounds between 40% and 50%. This skewness in the group uncertainty suggests high prevalence rates are believed to be unlikely but plausible.

[Fig. 1 about here.]

3.2 Pathway prevalence

The outputs of these models are the simulated posterior distributions for the model parameters, generated by the Gibbs sampler. Table 3 summarises the prior and posterior distributions in terms of medians and credible intervals for a model with and without covariances.

Figure 2 shows the smoothed densities of the posterior distributions of the prevalence of *P. psidii* s.l. for the three pathways, for the models with and without the assumption of conditional independence. The number of cycles was 20,000 so these represent the range of values observed for this set. These give an indication of the likely

prevalence values, given the prior information and data observed. The prevalence is estimated to be higher in containers than the other pathways in both cases, but the overall prevalence is generally estimated to be lower when we allow for correlations between tests. This is due to the fact that the correlation between the sensitivities of both of the tests, *covs*, was estimated to be the larger of the two in this model. For this reason, a positive test result for both Lab A and Lab B is no longer equivalent to two sources of independent evidence for the presence of *P. psidii* s.l., so we would estimate a lower prevalence.

In plant biosecurity, diagnostic tests are used to show disease presence and to assist in demonstrating freedom from disease of a host population. Although in this case the interest is in the freedom of a non-host plant exposure pathway from pathogenic spores, the principle is the same. For this reason, the probability of no spores being present (in biosecurity risk material) on a pathway in the surveyed period is a quantity of interest. An estimate of this probability can be obtained as the number of simulation cycles where the number of positives samples was zero. For example, for the flower samples using our model with conditional dependence, 8940 of the 20,000 cycles had zero positives ($Y_\bullet = 0$) so we estimate that there is a 44.7% chance of no spores being present.

[Table 3 about here.]

[Fig. 2 about here.]

3.3 Lab sensitivity and specificity

The main interest in this project was the estimation of the prevalence of *P. psidii* s.l. spores, but the outputs of the Gibbs sampler also allow assessment of the performance of the labs, based on the sensitivity and specificity values estimated from the marginal posteriors, which incorporate both the known sample and pathway sample data. Figure 3 shows the empirical cumulative density functions of the sensitivity and specificity for the flower data allowing for conditional dependence between tests, which are smooth curves

fitted to the simulation outputs. As these indicate, the test results from Lab B are more sensitive and more specific than those from Lab A. The main effect of allowing for the dependence between tests was to reduce the performance of Lab B, as well as reduce the specificity of Lab A. This is a predictable consequence of positive dependence between tests: some of the agreement between the tests, instead of providing additional evidence for test accuracy, can be explained by the correlation in the likelihood of a false positive or a false negative.

We considered combining sensitivity and specificity estimates across the three pathways, but the estimated posterior distributions for these quantities were found to be too heterogeneous between the pathways. That is, the combined simulated distributions demonstrated greater variability than any of the individual simulated distributions.

Generally, the performance of the tests on the pathway samples were poorer than the performance on the known samples, as evidenced by the lower sensitivity and specificity estimates in Table 3.

[Fig. 3 about here.]

Even with limited overlap in the samples assessed by both Lab A and Lab B, the estimates of the correlations between the two tests are non-negligible, and have been shown to impact the estimates of the key quantities of interest in terms of lower test performance. Therefore, the results that adjust for these correlations are deemed to be the most appropriate for interpretation.

3.4 Sensitivity of outputs to prior distributions

It is important at this point to consider how the estimated posterior distributions depend on the particular choices of prior distributions. Dependence on the prior distribution is not grounds for rejecting a posterior distribution – this is the aim of an informative prior – but this sensitivity ought to be assessed to properly interpret the results and the relative influence of the information used.

In order to assess the dependence of the model outputs on the prior distributions, the primary model outputs (generated by informative priors) were compared to equivalent outputs generated with uninformative (uniform) prior distributions. As the dependence model was deemed the most appropriate for this particular problem, results for the sensitivity analysis of this model have been presented.

Arguably, no distribution is strictly uninformative, as assigning equal probabilities to all outcomes is in itself informing the analysis (Dongen, 2006). The use of uniform priors here is to provide a comparison with the informative priors used in the primary analysis, for the purpose of sensitivity analysis. Since prior data were available in this case, ignoring these data and assuming uniform priors would not have been appropriate for the primary analysis.

As there were only three degrees of freedom and seven parameters to estimate for the model with correlated tests, at least four priors must be informative for the model to be identifiable (this compares to two in the conditionally independent case). See Section 4 for further discussion of the impact of this issue.

Table 4 provides the estimate for prevalence when uniform priors are used for that parameter for this model. This indicates negligible influence of the prior distributions on the prevalence, except for the case of containers, where the estimation had greater variance. There was also some evidence of sensitivity to the the prior distribution for the prevalence in the case of the passenger pathway. The expert judgement, as captured in this prior, was that the prevalence was almost certainly positive for the passenger pathway, resulting in larger posterior prevalence estimates than when using a uniform prior. Note that Table 4 does not include the sensitivity to the prior for the correlations, which are already uniform.

[Table 4 about here.]

[Table 5 about here.]

When uniform priors were used for the test sensitivity and specificity, substantial differences can be observed in the posterior distributions for these quantities,

as indicated in Table 5. In particular, the posterior distributions for test sensitivity in both Lab A and Lab B are closer to a uniform distribution. For example, the median of the posterior distribution for the sensitivity of Lab A in the passenger pathway was 0.63 (95% CI: 0.21,0.91) compared with 0.79 (95% CI: 0.62,0.91). This is evidence of strong dependence on the prior distributions for the estimates of test sensitivity, which were constructed from data obtained from tests of samples with known pathogen status. Without this measure of test accuracy from known samples, we would have little confidence in the reliability of the results for unknown pathway samples.

4 Discussion

4.1 Advantages and disadvantages of this approach

In epidemiological contexts, and in environmental management, information pertaining to risk and parameter estimation is often gradually accumulated from a variety of sources (Dunson, 2001; Pollino, White, and Hart, 2007). In these cases, there are a number of advantages to using Bayesian analysis, as summarised by Dunson (2001). Particular advantages include the ability of such methods to accommodate prior information and unobservable variables. The analysis also produces uncertainty estimates surrounding these quantities. The Bayesian MCMC approach, in particular, allows for a broad range of model specifications (Martin et al., 2005).

The Bayesian approach has been criticised most notably in Andersen (1997) because of the continued high dependency of the posterior on the prior distribution as the number of observations increases. In particular, both models are non-identifiable without sufficient genuinely informative prior distributions, as discussed by Dendukuri and Joseph (2001).

Joseph (1997) has responded to this critique with the argument that it was entirely appropriate that this be the case when $S < 1$ and/or $C < 1$ as the inaccuracy of the test captured in the prior information ought to be preserved in posterior

prevalence estimation. This does, of course, imply that caution needs to be taken in determining the priors. However, the alternative of assuming no prior knowledge and imposing 'uninformative' priors, such as uniform distributions for sensitivity and specificity, generates results that fail to reflect some legitimately available knowledge. The use of priors constructed from data – which was the case for the sensitivity and specificity priors in this study – has been shown to increase not only the precision of estimates, but also the accuracy (Morris, Vesk, and McCarthy, 2013).

Typically, the alternative to the Bayesian approach is to use maximum likelihood approaches (Poole et al., 1996; Lew and Levy, 1989) and it is true in this case that a likelihood could be constructed using the control data, but these methods rely on assumptions such as normality and rely on large sample approximations, so constraints would have to be imposed for such approaches. Data from an additional population can be used to overcome this limitation (Johnson, Gastwirth, and Pearson, 2001), but the use of prior information in the Bayesian approach eliminates the need for this. This approach allows prior information to be incorporated in a principled way, rather than in an ad hoc way which would be required by frequentist approaches. As the densities of Figures 2 and 3 indicate, it is not appropriate to assume normal distributions in each case.

A more detailed review of a range of approaches to diagnostic testing in the absence of a gold standard can be found in Reitsma et al. (2009).

4.2 Appropriateness of assumptions for pathway samples

It is important to note that the method used in this study was initially developed by Joseph et al. (1995) for different types of tests, with potentially very different accuracy. This is different to the pathway samples where the same PCR assay was applied at two different labs. Differences between labs (such as reagent suppliers, equipment manufacturers, etc.) are normal when an assay is implemented in different locations, though it is characteristic of a good assay to be robust to these differences to demonstrate repro-

ducibility. Note that it was inconsistencies between results of the two labs on pathway samples that initially indicated the potential fallibility of the testing process.

This study uses an approach that assumes conditional independence of the results for different labs and an extension that does not make this assumption. Toft et al. (2005) use a simulation study to highlight the potential impact of this assumption. As well as evidence from the data itself, indicated above, there are other reasons to suppose that there might be positive correlation between test results for a given sample, that is, the probability of the same outcome for each test might be greater than the sum of the marginal probabilities for each test. The assumption of conditional independence is difficult to satisfy fully in many similar cases, even where two different types of test are used (e.g., a PCR assay and a microscope inspection for spores). This is because, although the two tests will likely have greater differences in sensitivity and specificity than the intra-lab differences for the same PCR assay, the same underlying process that reduces sensitivity (for example) affects both types of test. That is, the degree of contamination of the sample by extraneous material collected during sampling influences both the sensitivity of the PCR assay (due to inhibition) and the effectiveness of microscope inspection (due to the reduction in visibility).

The method presented here also assumes that the missing results are missing at random; that is, that no additional knowledge of the true status of a sample can be gained from the fact that it has only been tested by one lab. While the allocation of the data to labs was not explicitly randomised, there is no reason to believe, for example, that the samples sent to Lab A only are different in nature to those samples sent to both labs.

The estimates for pathway prevalence are based on the assumption that the performance of the labs on the process control samples – for which the expected result is known – is representative of the performance of the labs on the pathway samples. This assumption may not be true because the known samples were produced by placing spores on clean pieces of filter paper, which does not completely simulate the matrix

of the pathway samples (which contain varying amounts of detritus). Results for dirty samples showed inhibition of the PCR that reduced the sensitivity of the test (data not shown); however, these dirty samples were also not precisely representative of the pathway sample matrix (which typically contained substantially less detritus than the tested dirty samples). So, while the dirty sample results indicate that inhibition could have been a factor in the pathway sample results, they do not constitute definitive evidence. It is, however, likely that it was a factor to at least some degree since inhibition of PCR is known to occur when applied to environmental samples (Stubner, 2002). This study showed how process controls can be used to improve test accuracy; creating process controls that more accurately represent the environmental samples being studied would improve this further.

4.3 Further extensions

This paper is a direct extension of methods proposed by Joseph et al. (1995) and Dendukuri and Joseph (2001). Other extensions to this paper also exist, including an extension to the case of three diagnostic tests (Scott et al., 2008) and the case of two populations with different prevalence values but the same diagnostic tests (Johnson et al., 2001). It is also possible to incorporate covariates into the estimation (Epstein, Munoz, and He, 1996; Tu, Kowalski, and Jia, 1999; Lewis, Sanchez-Vanquez, and Torgerson, 2012) and to estimate sample size for this and related methods (Beavers and Stamey, 2012). There may also be value in the use of second-order simulation to better understand the sources of uncertainty in the estimated posterior distributions (e.g., Nauta, 2000; Spiegelhalter and Best, 2003); although, also note criticism of common approaches to second-order simulation (Ferson and Ginzburg, 1996).

The two limitations discussed above, namely the assumption of conditional independence and tests missing at random, have been discussed elsewhere. In addition to the approach to address the former used here, there are alternatives that model the joint distribution between the two tests using, say, Dirichlet priors (Geisser

and Johnson, 1992; Mendoza-Blanco, Tu, and Iyengar, 1996; Black and Craig, 2002; Georgiadis, Johnson, Gardner, and Singh, 2003). When one diagnostic test is used to verify a previous positive test, which occurs frequently in medicine, the missing at random assumption is not appropriate and bias in the choice of test ought to be accounted for. This has been addressed using methods that estimate and adjust for verification bias (Lu et al., 2010; de Groot et al., 2012). There are also methods applicable where a gold standard is available for some samples but not all (Pennello, 2011).

The model proposed here assumes a non-zero prevalence. In the case of the application used here, this is equivalent to assuming there is some level of *P. psidii* s.l., although this level may be near zero. There are approaches which allow for zero prevalence estimates and which may be applicable in other related contexts; for example, Rahme and Joseph (1998). Other possible extensions are the use of hierarchical modelling to share information across pathways (Hanson et al., 2003) or two-step Bayesian modelling (Liu et al., 2014).

Some of these extensions are worth considering in the context of plant pathology in particular. Additional information available for each sample such as the amount of detritus present, the country of origin and other test results from the PCR process could be used to improve estimates of prevalence and test accuracy. This kind of information could be included via model covariates. The approach used in this study can also be extended to more than two tests, should results from an additional lab or alternative testing approach be made available. This would also overcome the potential for non-identifiability and reduce dependence on informative priors. Finally, it may be appropriate to explore other prior distributions for prevalence such as the beta-pert, which is more commonly used to represent expert opinion but which does not permit simple posterior distribution construction.

Acknowledgements

This study was supported by the Australian Centre of Excellence for Risk Analysis (now named the Centre of Excellence for Biosecurity Risk Analysis). We sincerely thank the experts who participated in the elicitation of priors, the biosecurity officers and volunteers who facilitated the collection of samples, and the scientists who performed the laboratory testing. We are also very grateful to Andrew Robinson, Bonnie Wintle, Marissa McBride and Mark Burgman for their advice. Finally, we appreciate the time taken by two anonymous reviewers to provide detailed comments which have greatly improved the manuscript.

APPENDIX A. Conditional distributions for Gibbs Sampler

APPENDIX A.1. Assuming conditional independence

Letting $Y_\bullet = \sum Y_i$,

$$Y_1 \mid n_1\pi, S_1, C_1, S_2, C_2 \sim \text{Binomial}\Big(n_1, \frac{\pi S_1 S_2}{\pi S_1 S_2 + (1-\pi)(1-C_1)(1-C_2)}\Big)$$

$$Y_2 \mid n_2\pi, S_1, C_1, S_2, C_2 \sim \text{Binomial}\Big(n_2, \frac{\pi S_1 (1-S_2)}{\pi S_1 (1-S_2) + (1-\pi)(1-C_1) C_2}\Big)$$

$$Y_3 \mid n_3\pi, S_1, C_1, S_2, C_2 \sim \text{Binomial}\Big(n_3, \frac{\pi (1-S_1) S_2}{\pi (1-S_1) S_2 + (1-\pi) C_1 (1-C_2)}\Big)$$

$$Y_4 \mid n_4\pi, S_1, C_1, S_2, C_2 \sim \text{Binomial}\Big(n_4, \frac{\pi (1-S_1)(1-S_2)}{\pi (1-S_1)(1-S_2) + (1-\pi) C_1 C_2}\Big)$$

$$Y_5 \mid n_5\pi, S_1, C_1 \sim \text{Binomial}\Big(n_5, \frac{\pi S_1}{\pi S_1 + (1-\pi_i)(1-C_1)}\Big)$$

$$Y_6 \mid n_6\pi, S_1, C_1 \sim \text{Binomial}\Big(n_6, \frac{\pi (1-S_1)}{\pi (1-S_1) + (1-\pi) C_1}\Big)$$

$$Y_7 \mid n_7\pi, S_2, C_2 \sim \text{Binomial}\Big(n_7, \frac{\pi S_2}{\pi S_2 + (1-\pi)(1-C_2)}\Big)$$

$$Y_8 \mid n_8\pi, S_2, C_2 \sim \text{Binomial}\Big(n_8, \frac{\pi (1-S_2)}{\pi (1-S_2) + (1-\pi) C_2}\Big)$$

$$\pi \mid N, Y_\bullet \sim \text{Beta}(Y_\bullet + \alpha_\pi, N - Y_\bullet + \beta_\pi)$$

$$S_1 \mid Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, \alpha_{S1}, \beta_{S1} \sim \text{Beta}(Y_1 + Y_2 + Y_5 + \alpha_{S1}, Y_3 + Y_4 + Y_6 + \beta_{S1})$$

$$S_2 \mid Y_1, Y_2, Y_3, Y_4, Y_7, Y_8, \alpha_{S_2}, \beta_{S_2} \sim \text{Beta}(Y_1 + Y_3 + Y_7 + \alpha_{S_2}, Y_2 + Y_4 + Y_8 + \beta_{S_2})$$

$$C_1 \mid Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, \alpha_{C_1}, \beta_{C_1} \sim \text{Beta}(n_3 + n_4 + n_6 - (Y_3 + Y_4 + Y_6) + \alpha_{C_1},$$
$$n_1 + n_2 + n_5 - (Y_1 + Y_2 + Y_5) + \beta_{C_1})$$

$$C_2 \mid Y_1, Y_2, Y_3, Y_4, Y_7, Y_8, \alpha_{C_2}, \beta_{C_2} \sim \text{Beta}(n_2 + n_4 + n_8 - Y_2 - Y_4 - Y_8 + \alpha_{C_2},$$
$$n_1 + n_3 + n_7 - Y_1 - Y_3 - Y_7 + \beta_{C_2})$$

APPENDIX A.2. Allowing for conditional independence

The following posterior distributions do not have a known form so sampling importance resampling (Rubin, 1988) was used to sample the values at each iteration. Letting $Y_\bullet = \sum Y_i$, $u_s = min(S_1, S_2) - S_1S_2$ and $u_c = min(C_1, C_2) - C_1C_2$,

$$
\begin{aligned}
Y_1 \mid n_1\pi, S_1, C_1, S_2, C_2, covs, covc &\sim \text{Binomial}\Big(n_1, \frac{\pi(S_1S_2 + covs)}{\pi(S_1S_2 + covs) + (1-\pi)((1-C_1)(1-C_2) + covc)}\Big)\\
Y_2 \mid n_2\pi, S_1, C_1, S_2, C_2, covs, covc &\sim \text{Binomial}\Big(n_2, \frac{\pi(S_1(1-S_2) - covs)}{\pi(S_1(1-S_2) - covs) + (1-\pi)((1-C_1)C_2 - covc)}\Big)\\
Y_3 \mid n_3\pi, S_1, C_1, S_2, C_2, covs, covc &\sim \text{Binomial}\Big(n_3, \frac{\pi((1-S_1)S_2 - covs)}{\pi((1-S_1)S_2 - covs) + (1-\pi)(C_1(1-C_2) - covc)}\Big)\\
Y_4 \mid n_4\pi, S_1, C_1, S_2, C_2, covs, covc &\sim \text{Binomial}\Big(n_4, \frac{\pi((1-S_1)(1-S_2) + covs)}{\pi((1-S_1)(1-S_2) + covs) + (1-\pi)(C_1C_2 + covc)}\Big)\\
Y_5 \mid n_5\pi, S_1, C_1 &\sim \text{Binomial}\Big(n_5, \frac{\pi S_1}{\pi S_1 + (1-\pi_i)(1-C_1)}\Big)\\
Y_6 \mid n_6\pi, S_1, C_1 &\sim \text{Binomial}\Big(n_6, \frac{\pi(1-S_1)}{\pi(1-S_1) + (1-\pi)C_1}\Big)\\
Y_7 \mid n_7\pi, S_2, C_2 &\sim \text{Binomial}\Big(n_7, \frac{\pi S_2}{\pi S_2 + (1-\pi)(1-C_2)}\Big)\\
Y_8 \mid n_8\pi, S_2, C_2 &\sim \text{Binomial}\Big(n_8, \frac{\pi(1-S_2)}{\pi(1-S_2) + (1-\pi)C_2}\Big)\\
\pi \mid N, Y_\bullet &\sim \text{Beta}(Y_\bullet + \alpha_\pi, N - Y_\bullet + \beta_\pi)
\end{aligned}
$$

$$
\begin{aligned}
p(S_1 \mid Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, S_2, \alpha_{S_1}, \beta_{S_1}, \beta_{covs}) &\propto (S_1S_2 + covs)^{Y_1}(S_1(1-S_2) - covs)^{Y_2}((1-S_1)S_2 - covs)^{Y_3}\\
&\times ((1-S_1)(1-S_2) + covs)^{Y_4}S_1^{Y_5}(1-S_1)^{Y_6}S_1^{\alpha_{S_1}-1}(1-S_1)^{\beta_{S_1}-1}(u_s - covs)^{\beta_{covs}-1}\\
p(S_2 \mid Y_1, Y_2, Y_3, Y_4, Y_7, Y_8, S_1, \alpha_{S_2}, \beta_{S_2}, \beta_{covs}) &\propto (S_1S_2 + covs)^{Y_1}(S_1(1-S_2) - covs)^{Y_2}((1-S_1)S_2 - covs)^{Y_3}\\
&\times ((1-S_1)(1-S_2) + covs)^{Y_4}S_2^{Y_7}(1-S_2)^{Y_8}S_2^{\alpha_{S_2}-1}(1-S_2)^{\beta_{S_2}-1}(u_s - covs)^{\beta_{covs}-1}\\
p(C_1 \mid Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, C_2, \alpha_{C_1}, \beta_{C_1}, \beta_{covc}) &\propto ((1-C_1)(1-C_2) + covc)^{n_1-Y_1}((1-C_1)C_2 - covc)^{n_2-Y_2}(C_1(1-C_2) - covc)^{n_3-Y_3}\\
&\times (C_1C_2 + covc)^{n_4-Y_4}(1-C_1)^{n_5-Y_5}C_1^{n_6-Y_6}C_1^{\alpha_{C_1}-1}(1-C_1)^{\beta_{C_1}-1}(u_c - covc)^{\beta_{covc}-1}\\
p(C_2 \mid Y_1, Y_2, Y_3, Y_4, Y_7, Y_8, C_1, \alpha_{C_2}, \beta_{C_2}, \beta_{covc}) &\propto ((1-C_1)(1-C_2) + covc)^{n_1-Y_1}((1-C_1)C_2 - covc)^{n_2-Y_2}(C_1(1-C_2) - covc)^{n_3-Y_3}\\
&\times (C_1C_2 + covc)^{n_4-Y_4}(1-C_2)^{n_7-Y_7}C_2^{n_8-Y_8}C_2^{\alpha_{C_2}-1}(1-C_2)^{\beta_{C_2}-1}(u_c - covc)^{\beta_{covc}-1}\\
p(covs \mid S_1, S_2, Y_1, Y_2, Y_3, Y_4, u_s, \alpha_{covs}, \beta_{covs}) &\propto (S_1S_2 + covs)^{Y_1}(S_1(1-S_2) - covs)^{Y_2}((1-S_1)S_2 - covs)^{Y_3}\\
&\times ((1-S_1)(1-S_2) + covs)^{Y_4}covs^{\alpha_{covs}-1}(u_s - covs)^{\beta_{covs}-1}\\
p(covs \mid C_1, C_2, Y_1, Y_2, Y_3, Y_4, u_c, \alpha_{covc}, \beta_{covc}) &\propto ((1-C_1)(1-C_2) + covc)^{n_1-Y_1}((1-C_1)C_2 - covc)^{n_2-Y_2}\\
&\times (C_1(1-C_2) - covc)^{n_3-Y_3}(C_1C_2 + covc)^{n_4-Y_4}covc^{\alpha_{covc}-1}(u_c - covc)^{\beta_{covc}-1}
\end{aligned}
$$

References

Andersen S (1997) Re: "Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard". American Journal of Epidemiology 145:290–291

Aven T (2013) On the meaning and use of the risk appetite concept. Risk Analysis 33:462–468

Beavers DP, Stamey JD (2012) Bayesian sample size determination for binary regression with a misclassified covariate and no gold standard. Computation Statistics and Data Analysis 56:2574–2582

Black MA, Craig BA (2002) Estimating disease prevalence in the absence of a gold standard. Statistics in Medicine 21:2653–2669

Burgman MA, McBride M, Ashton R, Speirs-Bridge A, Flander L, Wintle B, Fidler F, Rumpff L, Twardy C (2011) Expert status and performance. PLoS ONE 6:e22,998

Carnegie AJ, Lidbetter JR, Walker J, Horwood MA, Tesoriero L, Glen M, Priest MJ (2010) *Uredo rangelii*, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. Australasian Plant Pathology 39:463–466

Casella G, George EI (1992) Explaining the Gibbs Sampler. The American Statistician 46:167–174

Coutinho TA, Wingfield MJ, Alfenas AC, Crous PW (1998) Eucalyptus rust: a disease with the potential for serious international implications. Plant Disease 82:819–825

Demissie K, White N, Joseph L, Ernst P (1998) Bayesian estimation of asthma prevalence, and comparison of exercise and questionnaire diagnostics in the absence of a gold standard. Annals of Epidemiology 8:201–208

Dendukuri N, Joseph L (2001) Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 57:158–167

Dongen SV (2006) Prior specification in Bayesian statistics: three cautionary tales. Journal of Theoretical Biology 242:90–100

Dorny P, Phiri IK, Vercruysse J, Gabriel S, Willingham AL, 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

Dunson DB (2001) Commentary: practical advantages of Bayesian analysis of epidemiologic data. American Journal of Epidemiology 153:1222–1226

Enoe C, Georgiadis MP, Johnson WO (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

Epstein LD, Munoz A, He D (1996) Bayesian imputation of predictive values when covariate information is available and gold standard diagnosis is unavailable. Statistics in Medicine 15:463–476

Ferson S, Ginzburg LR (1996) Different methods are needed to propagate ignorance and variability. Reliability Engineering & System Safety 54(2):133–144

Garber A, Solomon N (1999) Cost-effectiveness of alternative test strategies for the diagnosis of coronary artery disease. Annals of Internal Medicine 130:719–728

Geisser S, Johnson W (1992) Optimal administration of dual screening tests for detecting a characteristic with special reference to low prevalence diseases. Biometrics 48:839–852

Georgiadis MP, Johnson WO, Gardner IA, Singh R (2003) Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. Journal of the Royal Statistical Society: Series C (Applied Statistics) 52(1):63–76

Glen M, Alfenas AC, Zauza EAV, Wingfield MJ, Mohammed C (2007) *Puccinia psidii*: a threat to the Australian environment and economy - a review. Australasian Plant Pathology 36:1–16

Grgurinovic CA, Walsh D, Macbeth F (2006) Eucalyptus rust caused by *Puccinia psidii* and the threat it poses to Australia. EPPO Bulletin 36:486–489

de Groot JAH, Dendukuri N, Janssen KJM, Reitsma JB, Brophy J, Joseph L, Bossuyt PMM, Moons KGM (2012) Adjusting for partial verification or workup bias in meta-

analyses of diagnostic accuracy studies. American Journal of Epidemiology 174:847–853

Hanson T, Johnson WO, Gardner IA (2003) Hierarchical models for estimating herd prevalence and test accuracy in the absence of a gold standard. Journal of agricultural, biological, and environmental statistics 8(2):223–239

Holliday JL, Jones SA, Simpson JA, Glen M, Edwards J, Robinson A, Burgman MA (2013) A novel spore collection device for sampling exposure pathways: a case study of *Puccinia psidii*. Plant Disease 97:828–834

Johnson WO, Gastwirth JL, Pearson LM (2001) Screening without a "Gold Standard": the Hui-Walter paradigm revisited. American Journal of Epidemiology 153:921–924

Joseph L (1997) Re: "Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard" The first author replies. American Journal of Epidemiology 145:291

Joseph L, Gyorkos TW, 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 141:263–272

Kawanishi T, Uematsu S, Kakishima M, Kagiwada S, Hamamoto H, Horie H, Namba S (2009) First report of rust disease on ohia and the causal fungus, *Puccinia psidii*, in Japan. Journal of General Plant Pathology 75:428–431

Keefer DL, Bodily SE (1983) Three-point approximations for continuous random variables. Management Science 29:595–609

Krull CR, Waipara NW, Choquenot D, Burns BR, Gormley AM, Stanley MC (2013) Absence of evidence is not evidence of absence: feral pigs as vectors of soil-borne pathogens. Austral Ecology 38:534–542

Langrell SRH, Glen M, Alfenas AC (2008) Molecular diagnosis of *Puccinia psidii* (guava rust) - a quarantine threat to Australian eucalypt and Myrtaceae biodiversity. Plant Pathology 57:687–701

Lew RA, Levy PS (1989) Estimation of prevalence on the basis of screening tests. Statistics in Medicine 8(10):1225–1230

Lewis F, Sanchez-Vanquez MJ, Torgerson PR (2012) Association between covariates and disease occurrence in the presence of diagnostic error. Epidemiology and Infection 140:1515–1524

Liu J, Chen F, Yu H, Zeng P, Liu L (2014) A two-stage Bayesian method for estimating accuracy and disease prevalence for two dependent dichotomous screening tests when the status of individuals who are negative on both tests is unverified. BMC medical research methodology 14(1):110

Lu Y, Dendukuri N, Schiller I, Joseph L (2010) A Bayesian approach to simultaneously adjusting for verification and reference standard bias in diagnostic test studies. Statistics in Medicine 29:2532–2543

Makowski D, Denis J-B, Ruck L, Penaud A (2008) A Bayesian approach to assess the accuracy of a diagnostic test based on plant disease measurement. Crop Protection 27:1187–1193

Martin TG, Kuhnert PM, Mengersen K, Possingham HP (2005) The power of expert opinion in ecological models using Bayesian methods: impact of grazing on birds. Journal of Applied Ecology 15:266–280

McBride MF, Fidler F, Burgman MA (2012) Evaluating the accuracy and calibration of expert predictions under uncertainty: predicting the outcomes of ecological research. Diversity and Distributions 18:782–794

Mendoza-Blanco JR, Tu XM, Iyengar S (1996) A Bayesian inference on prevalence using a missing-data approach with simulation-based techniques: applications to HIV screening. Statistis in Medicine 15:2161–2176

Mila AL, Yang XB, Carriquiry AL (2003) Bayesian logistic regression of soybean Sclerotinia stem rot prevalence in the U.S. north central region: accounting for uncertainty in parameter estimation. Phytopathology 93:758–764

Mireku E, Simpson JA (2002) Fungal and nematode threats to Australian forests and amenity trees from importation of wood and wood products. Canadian Journal of Plant Pathology 24:117–124

Morris WK, Vesk PA, McCarthy MA (2013) Profiting from pilot studies: analysing mortality using Bayesian models with informative priors. Basic and Applied Ecology 14:81–89

Nauta MJ (2000) Separation of uncertainty and variability in quantitative microbial risk assessment models. International Journal of Food Microbiology 57(1):9–18

Pennello GA (2011) Bayesian analysis of diagnostic test accuracy when disease state is unverified for some subjects. Journal of Biopharmaceutical Statistics 21:954–970

Pollino CA, White AK, Hart BT (2007) Examination of conflicts and improved strategies for the management of an endangered Eucalypt species using Bayesian networks. Ecological Modelling 201:37–59

Poole WK, Flynn PM, Rao AV, Cooley PC (1996) Estimating the prevalence of drug use from self-reports in a cohort for which biologic data are available for a subsample. American Journal of Epidemiology 144:413–420

Rahme E, Joseph L (1998) Estimating the prevalence of a rare disease: adjusted maximum likelihood. Journal of the Royal Statistical Society: Series D (The Statistician) 47(1):149–158

Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM (2009) A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. Journal of clinical epidemiology 62(8):797–806

Rubin DB (1988) Using the SIR algorithm to simulate posterior distributions. Bayesian statistics 3(1):395–402

Schiffman M, Wheeler CM, Dasgupta A, Solomon D, Castle PE (2005) A comparison of a prototype PCR assay and Hybrid Capture 2 for detection of carcinogenic human papillomavirus DNA in women with equivocal or mildly abnormal papanicolaou smears. American Journal of Clinical Pathology 124:722–732

Scott AN, Joseph L, Belisle P, Behr MA, Schwartzman K (2008) Bayesian modelling of tuberculosis clustering from DNA fingerprint data. Statistics in Medicine 27:140–156

Simpson JA, Thomas K, Grgurinovic CA (2006) Uredinales species pathogenic on species of Myrtaceae. Australasian Plant Pathololology 35:549–562

Speirs-Bridge A, Fidler F, McBride M, Flander L, Cumming G, Burgman M (2010) Reducing overconfidence in the interval judgments of experts. Risk Analysis 30:512–523

Spiegelhalter DJ, Best NG (2003) Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Statistics in Medicine 22(23):3687–3709

Stubner S (2002) Enumeration of 16S rDNA of Desulfotomaculum lineage 1 in rice field soil by real-time PCR with SybrGreen$_{\mathrm{TM}}$ detection. Journal of Microbiological Methods 50:155–164

Toft N, Jørgensen E, Højsgaard 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(1):19–33

Tommerup IC, Alfenas AC, Old KM (2003) Guava rust in Brazil - a threat to eucalyptus and other Myrtaceae. New Zealand Journal of Science 33:420–428

Tu XM, Kowalski J, Jia G (1999) Bayesian analysis of prevalence with covariates using simulation-based techniques: applications to HIV screening. Statistics in Medicine 18:3059–3073

Vettraino AM, Sukno S, Vannini A, Garbelotto M (2010) Diagnostic sensitivity and specificity of different methods used by two laboratories for the detection of *Phytophthora ramorum* on multiple natural hosts. Plant Pathology 59:289–300

Vidal E, Moreno A, Berolini E, Cambra M (2012) Estimation of the accuracy of two diagnostic methods for the detection of Plum pox virus in nursery blocks by latent class models. Plant Pathology 61:413–422

Vijan S, Hwang EW, Hofer TP, Hayward RA (2001) Which colon cancer screening test? a comparison of costs, effectiveness, and compliance. American Journal of Medicine 111:593–601

Vlems FA, Ladanyi A, Gertler R, Rosenberg R, Diepstra JHS, Rder C, Nekarda H, Molnar B, Tulassay Z, van Muijen GNP, Vogel I (2003) Reliability of quantitative reverse-transcriptase-PCR-based detection of tumour cells in the blood between different laboratories using a standardised protocol. European Journal of Cancer 39:388396

List of Figures

1 29
2 30
3 31

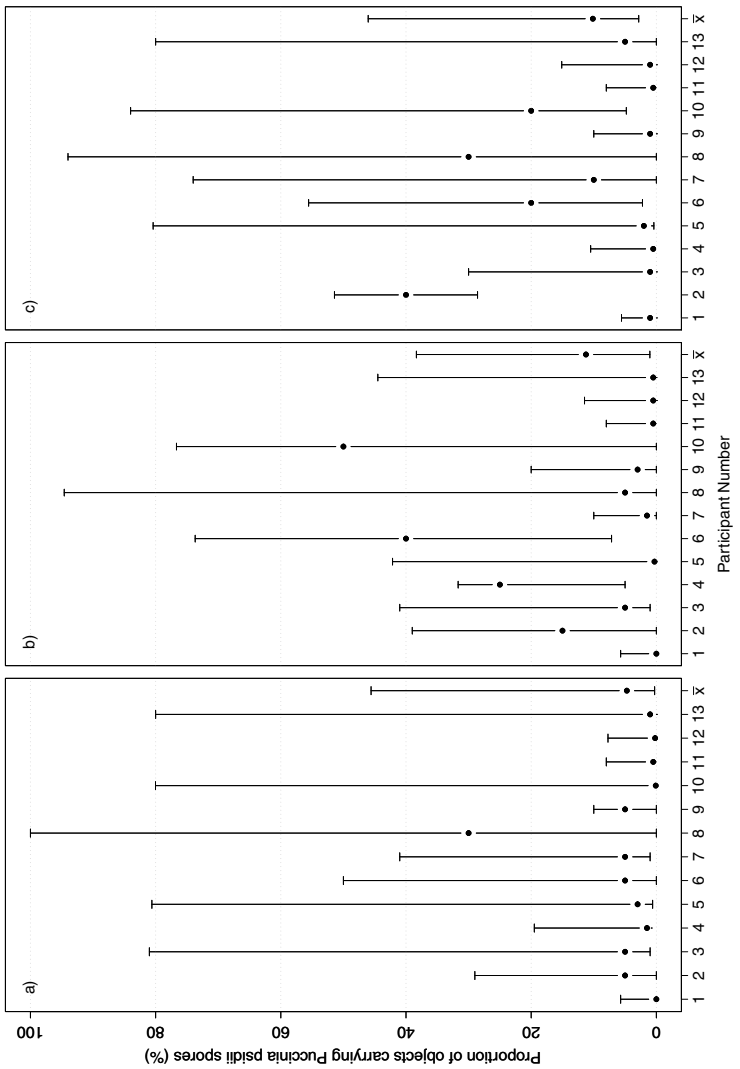

Fig. 1 Results of expert elicitation to estimate the prevalence of *Puccinia psidii* s.l. spores on pathways: a) flowers, b) passengers, and c) containers. The points indicate the participants' best estimates, while the upper and lower bars indicate the participants' standardised 80% confidence intervals (CIs). The average ($\bar{x}$) is the unweighted group average.

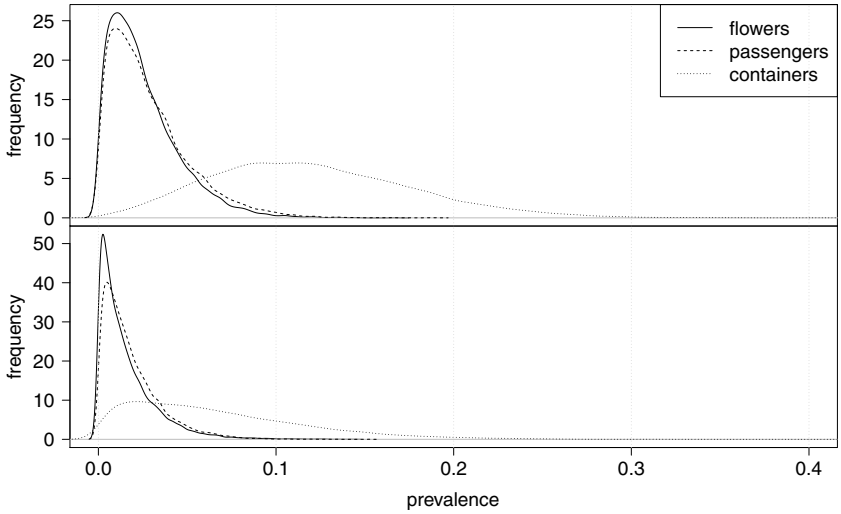

Fig. 2 Prevalence outputs from the Gibbs sampler for flowers (top), passengers (middle) and containers (bottom), i) assuming conditional independence between tests and ii) adjusting for correlations between tests.

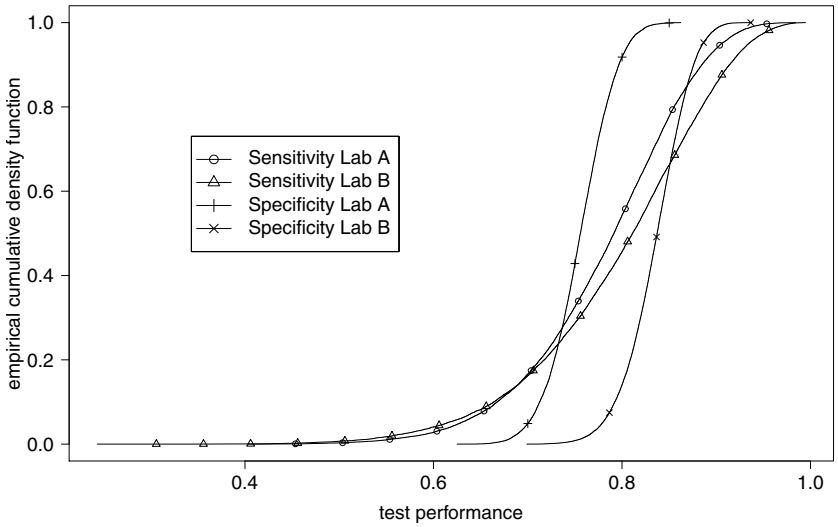

Fig. 3 Cumulative density function for sensitivity and specificity based on flower data.

List of Tables

1 Observed data from individual diagnostic tests for each pathway, where each test is either positive (+), negative (−) or not present (NA). The actual total sample sizes (N) for each pathway were 18, 54 and 27 for flowers, passengers and containers, respectively. 33
2 Likelihood contributions of all possible combinations of observed and latent data for the cases of two diagnostic tests for each pathway, including the additional contribution of the covariances for the model with correlated tests. The likelihood is proportional to the product of each entry in the last column raised to the power of the responding entry in the first column of the table. 34
3 Summary of the marginal prior and posterior distributions in terms of credible intervals. *The credible intervals for π are actually 80% intervals. 35
4 Summary of the marginal posterior distribution of prevalence with informative and uniform priors, to understand the impact of the prior distributions. The first row for each pathway contains the results based on all informative priors, followed by the results with uniform priors for each parameter. 36
5 Summary of the marginal posterior distributions of test sensitivity and specificity with informative and uniform priors, to understand the impact of the prior distributions. 37

		Test 2			
		+	−	NA	
Test 1	+	n_1	n_2	n_5	
	−	n_3	n_4	n_6	
	NA	n_7	n_8	-	
					N

Table 1 Observed data from individual diagnostic tests for each pathway, where each test is either positive (+), negative (−) or not present (NA). The actual total sample sizes (N) for each pathway were 18, 54 and 27 for flowers, passengers and containers, respectively.

No. of subjects	Truth	Test 1 result	Test 2 result	Likelihood contribution	Covariance contribution
Y_1	+	+	+	$\pi S_1 S_2$	$+covs$
Y_2	+	+	−	$\pi S_1(1-S_2)$	$-covs$
Y_3	+	−	+	$\pi(1-S_1)S_2$	$-covs$
Y_4	+	−	−	$\pi(1-S_1)(1-S_2)$	$+covs$
Y_5	+	+	NA	πS_1	
Y_6	+	−	NA	$\pi(1-S_1)$	
Y_7	+	NA	+	πS_2	
Y_8	+	NA	−	$\pi(1-S_2)$	
n_1-Y_1	−	+	+	$(1-\pi)(1-C_1)(1-C_2)$	$+covc$
n_2-Y_2	−	+	−	$(1-\pi)(1-C_1)C_2$	$-covc$
n_3-Y_3	−	−	+	$(1-\pi)C_1(1-C_2)$	$-covc$
n_4-Y_4	−	−	−	$(1-\pi)C_1C_2$	$+covc$
n_5-Y_5	−	+	NA	$(1-\pi)(1-C_1)$	
n_6-Y_6	−	−	NA	$(1-\pi)C_1$	
n_7-Y_7	−	NA	+	$(1-\pi)(1-C_2)$	
n_8-Y_8	−	NA	−	$(1-\pi)C_2$	

Table 2 Likelihood contributions of all possible combinations of observed and latent data for the cases of two diagnostic tests for each pathway, including the additional contribution of the covariances for the model with correlated tests. The likelihood is proportional to the product of each entry in the last column raised to the power of the responding entry in the first column of the table.

pathway		Prior information		Posterior (independence)		Posterior (correlated tests)	
		Median	95% CI	Median	95% CI	Median	95% CI
flowers	$\pi*$	0.05	0.00-0.41	0.02	0.00-0.06	0.01	0.00-0.05
	S_1	0.87	0.60-0.98	0.89	0.72-0.98	0.79	0.63-0.91
	C_1	0.83	0.74-0.89	0.77	0.72-0.83	0.76	0.70-0.81
	S_2	1.00	0.63-1.00	0.97	0.79-1.00	0.81	0.62-0.94
	C_2	0.99	0.93-1.00	0.99	0.96-1.00	0.84	0.78-0.89
	$covs$					0.05	0.01-0.14
	$covc$					0.04	0.00-0.10
passengers	$\pi*$	0.11	0.01-0.38	0.02	0.00-0.07	0.01	0.00-0.05
	S_1	0.87	0.60-0.98	0.86	0.66-0.97	0.79	0.63-0.91
	C_1	0.83	0.74-0.89	0.74	0.68-0.79	0.73	0.67-0.78
	S_2	1.00	0.63-1.00	0.97	0.79-1.00	0.81	0.61-0.93
	C_2	0.99	0.93-1.00	0.96	0.92-0.99	0.84	0.79-0.88
	$covs$					0.05	0.00-0.14
	$covc$					0.01	0.00-0.04
containers	$\pi*$	0.10	0.03-0.46	0.11	0.04-0.22	0.06	0.01-0.17
	S_1	0.87	0.60-0.98	0.84	0.66-0.96	0.80	0.64-0.91
	C_1	0.83	0.74-0.89	0.77	0.71-0.83	0.75	0.69-0.81
	S_2	1.00	0.63-1.00	0.97	0.81-1.00	0.81	0.60-0.94
	C_2	0.99	0.93-1.00	0.98	0.93-1.00	0.82	0.76-0.87
	$covs$					0.05	0.00-0.13
	$covc$					0.02	0.00-0.07

Table 3 Summary of the marginal prior and posterior distributions in terms of credible intervals. *The credible intervals for π are actually 80% intervals.

pathway	uniform parameter	Median	95% CI
flowers	-	0.011	0.001-0.047
	π	0.012	0.001-0.051
	S_1	0.010	0.001-0.044
	C_1	0.009	0.000-0.040
	S_2	0.012	0.001-0.053
	C_2	0.010	0.001-0.047
passengers	-	0.014	0.001-0.050
	π	0.011	0.001-0.045
	S_1	0.014	0.001-0.049
	C_1	0.012	0.001-0.043
	S_2	0.017	0.002-0.066
	C_2	0.014	0.001-0.051
containers	-	0.056	0.007-0.161
	π	0.045	0.004-0.147
	S_1	0.050	0.006-0.156
	C_1	0.032	0.003-0.111
	S_2	0.070	0.009-0.204
	C_2	0.055	0.007-0.160

Table 4 Summary of the marginal posterior distribution of prevalence with informative and uniform priors, to understand the impact of the prior distributions. The first row for each pathway contains the results based on all informative priors, followed by the results with uniform priors for each parameter.

pathway		informative		non-informative	
		Median	95% CI	Median	95% CI
flowers	S_1	0.79	0.63-0.91	0.64	0.22-0.92
	C_1	0.76	0.70-0.81	0.68	0.59-0.76
	S_2	0.81	0.61-0.93	0.59	0.17-0.89
	C_2	0.84	0.78-0.89	0.84	0.78-0.89
passengers	S_1	0.79	0.62-0.91	0.63	0.21-0.91
	C_1	0.73	0.68-0.78	0.63	0.56-0.71
	S_2	0.81	0.60-0.93	0.54	0.14-0.88
	C_2	0.84	0.79-0.88	0.84	0.79-0.88
containers	S_1	0.80	0.64-0.91	0.70	0.27-0.93
	C_1	0.75	0.69-0.81	0.57	0.45-0.69
	S_2	0.81	0.61-0.93	0.54	0.15-0.88
	C_2	0.82	0.76-0.87	0.82	0.76-0.87

Table 5 Summary of the marginal posterior distributions of test sensitivity and specificity with informative and uniform priors, to understand the impact of the prior distributions.