Te Rangahau Aroturuki i ngā Rākau Rangatira o Te Wao Nui ā Tiriwa

2021 Waitākere Ranges Kauri Population Health Monitoring Survey

June 2022, Technical Report 2022/8

aucklandcouncil.govt.nz

Chapter 4

Estimation of the diagnostic sensitivity and specificity of kauri dieback visual assessment and *Phytophthora agathidicida* soil baiting, culturing and morphological identification using Bayesian latent class analysis

Te whakatau tatahanga o te aromatawai ā-tirohanga e ine ana i te tino putanga me te tino korenga o te puruheka patu kauri, te rumaki hoki i te one hei whakatipu i te puruheka patu kauri, hei whakarea hoki i taua puruheka rā, hei tautuhi hoki i te hanga mā tā Bayesian tātari i te momo e torohū ana

Authors

Emilie Vallee[1], Karyn Froud[2], John Kean[3], Yue Chin Chew[4], Cord Heuer[1], Lisa Tolich[4]

[1]Massey University

[2]Biosecurity Research

[3]AgResearch

[4]Auckland Council

4.1 Abstract

Te whakatūporotanga

An accurate and precise estimation of *Phytophthora agathidicida* diagnostic tests' performance is needed to design and interpret past and future surveillance work, including the identification of areas free of the pathogen. The tests are: i) an indirect test of visual assessment of trees for symptoms consistent with kauri dieback to predict or extrapolate the presence of *P. agathidicida* in association with a kauri tree and ii) a soil sampling, baiting, culture and morphological identification (referred to as the soil sampling bioassay) to detect *P. agathidicida* in association with a kauri tree. Test performance is measured by the diagnostic sensitivity (the probability of a tree that does have *P. agathidicida* in its soil returning a positive test result) and diagnostic specificity (the probability of a tree that does not have *P. agathidicida* in its soil returning a negative test result) not to be confused with analytical sensitivity and specificity more commonly discussed in plant pathology (refer to terminology). In the absence of a gold standard (perfect) test to determine the true *P. agathidicida* status of a kauri tree, Bayesian latent class analysis (BLCA) is used as a reference method to estimate the tests' performances.

A BLCA model was built using prior expert opinion on the tests' performance and pathogen prevalence where appropriate, and data was collected from 761 trees using visual assessment and the soil sampling bioassay. In total, 159 trees were sampled and visually assessed from an area that was delimited as having a high prevalence by experts and 572 trees from a low prevalence area in the Waitākere Ranges between March and July 2021. The two tests were assumed to be conditionally independent, which means that for a given true infection status, the probability of a given result for one was independent of the other test's result.

For visual assessment, the estimated sensitivity was 41.0% (95% PI 29.8-53.3) and the estimated specificity was 87.0% (95% PI 84.0-89.8).

For the soil sampling bioassay, the estimated sensitivity was 63.2% (95% PI 42.6-88.1) and the estimated specificity was 98.7% (95% PI 96.8-99.8). If we assumed a perfect specificity, i.e., if we assumed it could never give a false-positive result, which is reasonable for a culture test, the sensitivity remains similar at 63.8% (95% PI 43.3-89.1).

Limitations on these results included the fact that the priors were designed using expert elicitation on modifications of the tests (an 8-point rather than 4-point pooled sample) and a low sample size especially in the high prevalence area, leading to large credible intervals for sensitivity estimates.

Using the estimates from the present study to interpret previous surveillance work that used visual assessment and the soil sampling bioassay sequentially, it is likely that the true number of trees with *P. agathidicida* present is around 3.9 times what has historically been recorded.

Finally, the value of sensitivity for the soil sampling bioassay can be used to calculate sample sizes for the definition of areas free of *P. agathidicida*, which can be done easily if we assume a specificity of 100%. For example, if a sample size of 463 trees all test negative in an area with 10,000 kauri trees, we can be 95% confident that if *P. agathidicida* is present, it will be below a prevalence of 1%.

4.2 Introduction

Te whakataki

This study evaluates the diagnostic test performance of two tests that are used in surveillance to estimate the presence of *P. agathidicida* in soils beneath monitored kauri. The two tests are: i) an indirect test of visual assessment of trees for symptoms consistent with kauri dieback to predict or extrapolate the presence of *P. agathidicida* in association with a kauri tree and ii) a soil sampling, baiting, culture and morphological identification (referred to as the soil sampling bioassay) to detect *P. agathidicida* in association with a kauri tree. Obtaining accurate and precise estimates of diagnostic sensitivity (the probability of a truly positive individual to give a positive test result) and specificity (the probability of a truly negative individual to give a negative test result) of the tests used for monitoring is crucial to design and interpret the results of surveillance activities, including those previously completed. Diagnostic sensitivity and specificity refer to the performance of the full methods for a diagnostic test in a population (World Organisation for Animal Health, 2019, Cardwell et al., 2018). In this study we want to know how good our tests (visual assessment and soil sampling) are at diagnosing whether *P. agathidicida* is present or absent. Diagnostic sensitivity and sensitivity differ from, and can be confused with, analytical sensitivity and specificity, which are more commonly calculated for plant pathogen tests. Analytical sensitivity refers to the lowest level of target agent that can be measured accurately by the test (Cardwell et al., 2018) whereas analytical specificity is similar to diagnostic specificity but is concerned with performance around excluding non-target species and cross-reactions (false positives) in the laboratory (Cardwell et al., 2018). Traditionally, the estimation of diagnostic sensitivity and specificity directly follow the estimation of analytical sensitivity and specificity in the development and validation of diagnostic tests (Cardwell et al., 2018).

The diagnostic values are necessary for calculation of true prevalence estimates or sample sizes required to assign a site as *P. agathidicida*-free for management purposes (such as high-value protected areas). The values also allow land managers to compare tests so that the test (or tests) with the best characteristics for the surveillance question can be used. For example, tests with a high sensitivity are suitable for screening for a causal pathogen, and tests with a high specificity are useful for confirming disease caused by a specific pathogen (Dohoo et al., 2009). Possibly because of different disease surveillance designs and control goals, diagnostic sensitivity and

specificity have rarely been estimated for tests for plant diseases. The only New Zealand example is (Heuer and Taylor, 2015) who estimated diagnostic sensitivity and specificity for *Pseudomonas synringae* pv. *actinidae* PCR assays in kiwifruit and used the values to provide recommendations to interpret test results and design detection surveys.

The presence of kauri dieback symptoms is assessed visually, aerially (Jamieson et al., 2014) and/or on the ground, or using remote sensing (Meiforth, 2020, Meiforth et al., 2020). The symptoms can resemble manifestations of stress for other reasons. Accurate detection of symptoms and attribution to *P. agathidicida* as opposed to another cause of stress is likely dependent on the observer's experience and knowledge of the location. The visual assessment usually includes an inspection on the ground by trained surveyors, who in addition to checking symptoms, decide if they are compatible with kauri dieback and not just ill-thrift. A five-point scale of disease severity of the canopy ranging from 1 for healthy trees to 5 for dead trees has been created by Dick and Bellgard (2010). Visual assessment is quick and relatively easy for trained observers to use as a test, however, it is uncertain how well visual assessment can predict **presence** of *P. agathidicida* and indicate infection by *P. agathidicida.*

The presence or absence of *P. agathidicida* for surveillance purposes is currently mainly investigated using the soil baiting, culture and morphological assessment described by (Beever et al., 2010). The performance of the assay itself is likely to be dependent on the soil sampling protocol used, and high inter-laboratory variation has been observed in the past (Froud, 2020), but efforts to standardise testing have been made (Beauchamp, 2016, Kauri Dieback Programme, 2017). However, any estimation of sensitivity and specificity will be specific to the sampling protocol and laboratory used to provide the data. Current surveillance activities use either a four (Auckland Council) or eight (Department of Conservation, Ministry for Primary Industries) cardinal points sampling protocol, and samples are tested at one or two of three approved research laboratories. The soil sampling bioassay is relatively expensive, causes direct disturbance to kauri roots and it is uncertain how well it can confirm the **absence** of *P. agathidicida.*

Traditional methods to estimate diagnostic sensitivity and specificity require the use of a gold standard, which is defined as a perfect test that never gives false-negative and false-positive results. In most cases, however, such a test does not exist. Bayesian latent class analysis for diagnostic test evaluation in the absence of a gold standard (Johnson et al., 2019, Cheung et al., 2021) allows estimation of test sensitivity and specificity even when no perfect test is available for comparison. This report estimates the diagnostic sensitivity and diagnostic specificity for *P. agathidicida* detection of the kauri dieback visual assessment test and the soil sampling bioassay (soil sampling, baiting, culture and morphological identification) used by Auckland Council using Bayesian latent class analysis. Additionally, it provides true prevalence estimates for two sets of sampling areas in Te Wao Nui ā Tiriwa / the Waitākere Ranges, North Island, New Zealand. This report follows the STARD-BLCM (Standards for the Reporting of Diagnostic accuracy studies by the use of Bayesian Latent Class Models) reporting guidelines (Kostoulas et al., 2017).

4.3 Objectives

Ngā whāinga

The objectives of this work were to undertake diagnostic test performance evaluation using latent class models of the following two tests:

i. Visual assessment of trees to detect symptoms of disease against a case definition
ii. Soil sampling bioassay involving baiting, culturing, and morphological identification

4.4 Methods

Ngā tikanga

This study closely followed the protocol detailed in (Vallee et al., 2019), with some modifications as detailed in this section. This study uses a latent class analysis methodology, described below. The following assumptions were made and deemed reasonable:

- The diagnostic sensitivity and specificity of both tests are constant across the different areas and trees sampled
- The two tests are conditionally independent, which means that for a given true infection status, knowing the result of one test would not change the chance of the other test to return a positive result
- The high and low prevalence areas have prevalence different from each other, and different from 0% and 100%. In other words, both areas have truly infected and truly healthy trees.

4.4.1 Data

The diagnostic test evaluation was done retrospectively using data previously collected from the cross-sectional prevalence study described in Chapter 2.

4.4.2 Tree selection

Trees were selected independently of disease status in the Waitākere Ranges (Figure 4-1) as described in Chapter 2. High and low prevalence sites were informed by previous surveillance activities. Possible high prevalence areas were assessed by Alastair Jamieson (Auckland Council), a kauri dieback aerial surveillance expert very familiar with the Waitākere Ranges, who used knowledge gained from two rounds of aerial surveillance looking for canopy ill-thrift in the Waitākere Ranges to inform risk-based ground surveillance in 2012 and 2016 (Hill et al., 2017, Jamieson, 2012b). Areas were identified on a map as apparently high prevalence polygons, including the surrounding contiguous area that was considered likely also to be affected, with all

other areas considered low prevalence (Figure 4-1). These identified areas were cross-checked by local mana whenua who hold mātauranga Māori (cultural knowledge) of the health status of the forest. In total, 189 kauri from the predefined high prevalence area and 572 trees from the low prevalence area were randomly selected.

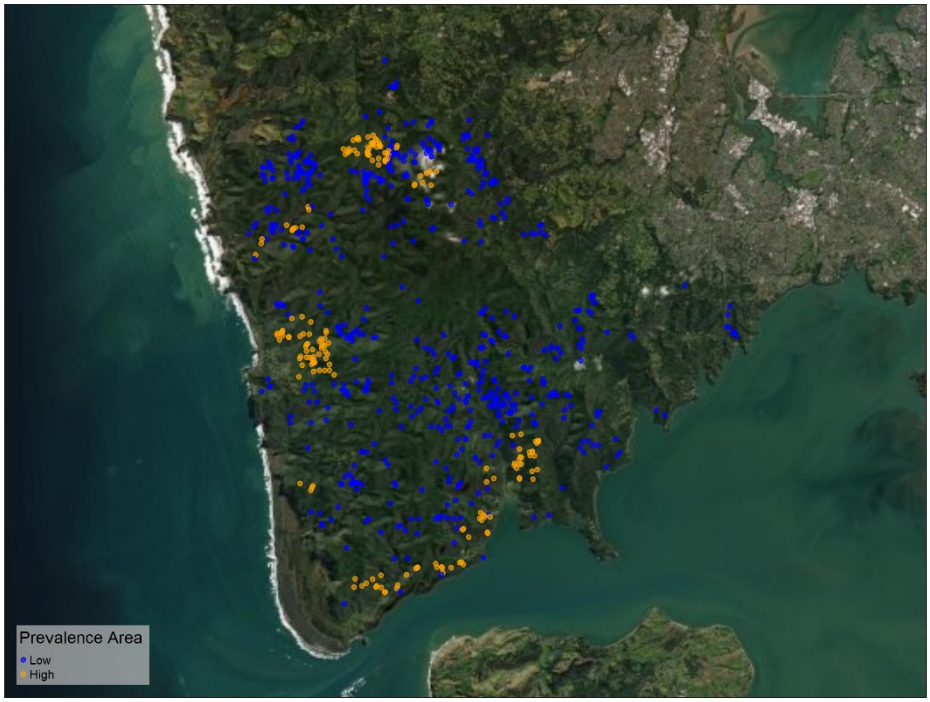


Figure 4-1. Locations of trees sampled in the Waitākere Ranges, North Island, New Zealand, for the evaluation of 2 kauri dieback diagnostic tests. Dots of tree locations from estimated low prevalence areas are in blue and dots for tree locations in estimated high prevalence areas are in yellow.

4.4.3 Visual assessment

Each pre-selected tree was visually assessed on the ground as described in Chapter 2 and using the case definition by Stevenson and Froud (2020). Surveyors observed the trees for the following symptoms: bleeding lesions on the basal trunk or lateral roots, the presence of canopy thinning (canopy score of 3 or higher as defined by Dick and Bellgard (2010)), yellowing of the foliage or copper-brown colour or tree death. Surveyors also observed the tree's surroundings to decide whether the observed symptoms were consistent with kauri dieback or could be attributed to another cause. Symptomatic trees, classified positive by visual assessment, were those showing at least one of the listed symptoms and where the surveyor decided they were consistent with possible or severe kauri dieback.

4.4.4 Soil sampling bioassay

Soil samples were collected around the base of pre-selected trees using the 4-cardinal point protocol at the time of visual assessment. Briefly, four samples were collected and pooled per tree and sent to Plant and Food Research, Havelock North, North Island, New Zealand for the soil bioassay which is described in Chapter 2.

4.4.5 Prior distributions for tests sensitivity and specificity, and prevalence

The method used for this analysis, based on Bayesian analyses, needs "prior" information that is credible, scientifically relevant, and formulated as probability distributions. These prior distributions reflect the knowledge of test performance and prevalence in the study area before this analysis, from recent studies and expert opinion. The priors used for the soil sampling bioassay (SB) were based on those obtained by (Vallee et al., 2019) using a formal expert elicitation process. The elicitation process followed the method described in Hemming et al. (2018): briefly, eight experts involved in *P. agathidicida* testing and kauri dieback management answered two rounds of an online survey asking for their opinion on the minimum, maximum and most likely value of the diagnostic sensitivity and specificity of the soil bioassay (with a modification of the sampling protocol, using 8 sampling points). Experts discussed the results of the first round face-to-face before doing the second round. Since the plant health discipline does not routinely use the diagnostic sensitivity and specificity concepts and because of the small change in sampling protocol, the intervals were modified before conducting the analysis to increase uncertainty and give more weight to the data. While the model used is identifiable (see 4.4.6 Model below), meaning that estimates of test performance can be obtained from the data only without the need for priors, the priors for SB were considered useful to help improve the precision of the posterior estimates.

"Flat" priors, giving an equal probability for all values between 0 and 100%, were used for the visual assessment (VA), as no reliable information was available. The use of these flat priors ensured that the values of sensitivity and specificity for VA were estimated only from the data, since the model was identifiable (See 4.4.6 Model below). The corresponding beta distribution is beta(1, 1).

Priors for high and low prevalence areas (pi1 and pi2 respectively) were based on previous aerial surveillance and expert opinion. Distributions were generated using BetaBuster, a purpose-built GUI to obtain Beta prior distributions (Su et al., 2012).

They are summarised in Table 4-1 and the distributions can be seen in the figures in Table 4-2 as well as Figure 4-3, Figure 4-5, Figure 4-7, and Figure 4-8.

Table 4-1. Prior belief and corresponding beta distributions for the different parameters needed to estimate the sensitivity and specificity of 2 tests for kauri dieback using BLCA

Parameter name and description	Description of prior belief	Prior distribution	Source
Se_{SB} (sensitivity of SB)	Min 65%, most likely 73% obtained from experts' elicitation; it was assumed this represented a 50% confidence interval	beta(2.89, 1.70)	Modified from Vallee et al. (2019)
Sp_{SB} (specificity of SB)	Min 86%, most likely 92% obtained from experts' elicitation;	beta(8.56, 1.66)	Modified from Vallee et al. (2019)

	it was assumed this represented a 50% confidence interval		
pi1 (high prevalence)	"40%, with some areas at 80%", set as 90% sure than lower than 80% and most likely at 50%	beta(2.06, 2.06)	A. Jamieson, Auckland Council, pers. comm.
pi2 (low prevalence)	50% sure that less than 5%, most likely at 4%	beta(3.42, 59.19)	A. Jamieson, Auckland Council, pers. comm.

The prior for SB specificity was narrower than for sensitivity, indicating that the experts had more confidence in their belief of specificity.

4.4.6 Model

The analysis follows the "two tests, two populations" method described in Branscum et al. (2005) and Johnson et al. (2019) and originally by Hui and Walter (1980). Briefly, the latent class analysis method used here relies on the existence of the true infection status, here the presence of *P. agathidicida* in the soil around a tree, that is unknown (latent) and that the two tests are measuring. It is the reference method to estimate a test's diagnostic sensitivity and specificity in the absence of a perfect, gold standard test and is recognised as such by the World Organisation for Animal Health (World Organisation for Animal Health, 2019). New developments and applications are regularly available (for example, see Cheung et al. (2021)).

The prior information on the parameters listed in Table 4-1, was combined with the data obtained from the tree visual assessment and the soil sampling bioassay in the two populations (Table 4-3) via a likelihood function representing the probability of observing the test results obtained after the tests were conducted as a function of the unknown parameters (sensitivities, specificities and prevalence).

Bayesian estimates of the 2.5th, 50th and 97.5th percentiles of the posterior probability distribution of Se_{VA}, Sp_{VA}, Se_{SB}, Sp_{SB}, high prevalence, low prevalence, the "inference after observing the data" (Johnson et al 2019), were then obtained using Markov Chain Monte Carlo (MCMC) chains with a Gibbs sampler with 50,000 iterations, and the first 10,000 were discarded for results presentations, to avoid any influence of values obtained before model convergence. Three chains were run in parallel, with spread initial values, and convergence was visually assessed on "trace" plots. The Gelman and Rubin's convergence diagnostics and the Gelman-Rubin-Brooks plot are presented in Appendix E. For more information on the Bayesian Latent Class Analysis method please refer to Branscum et al. (2005). The interpretation of the uncertainty intervals (named probability intervals PI) is more intuitive than the confidence intervals generated in a traditional frequentist (non-Bayesian) statistical approach. In other words, the 95% credible intervals presented in the results correspond to the 2.5th and 97.5th percentiles of the total number of iterations of the model. The analysis was conducted in

OpenBUGS (version 3.2.3, OpenBUGS Project Management Group, 2014). More details on the model structure and specification are found in the code in Appendix E.

4.4.7 Sensitivity analysis

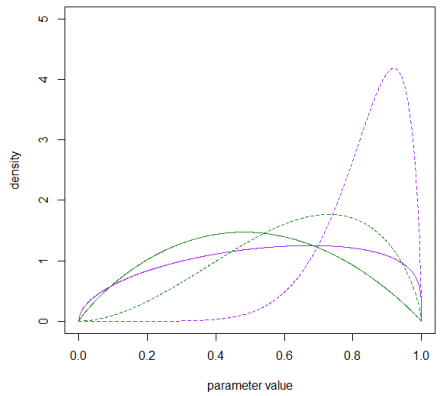
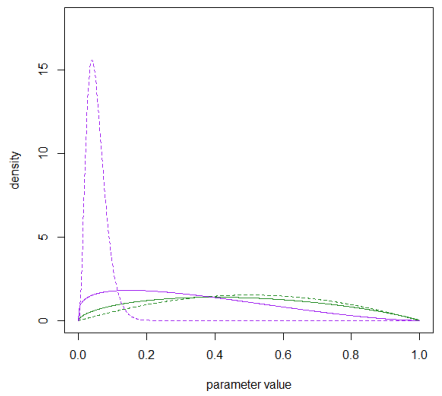
In the model used for this study, we are estimating 6 parameters (Se_{VA}, Sp_{VA}, Se_{SB}, Sp_{SB}, high prevalence pi1, low prevalence pi2). Under the assumption of conditional independence, the model is identifiable, which means that the values could be estimated using the data only, without the priors. Using priors is however often helpful, as it can help increasing the precision of the estimates. It is however important to assess the effect of the priors on the results to understand how they contribute to the final estimate and what effect any misspecification of the priors would have on the posterior distributions of the parameters. To assess the effect of priors on the results, the analysis was repeated seven times, each time with a small, plausible change in the prior distributions. The effect of the priors of VA, SB and prevalence were assessed separately, keeping the others constant. The prior distributions for test performance were obtained by modifying slightly the intervals given by the experts (see Vallee et al. (2019)) and transforming them into beta distribution using the 'prevalence' package in R that implements the method described by (Branscum et al., 2005), assuming an expert confidence of 80%. The priors for prevalence were obtained using BetaBuster (Su et al., 2012).

The following 7 changes to the prior distributions (Table 4-2) were used, in different runs of the model:

- Model run 1: The specificity for SB was fixed to 100%, with no uncertainty. Hence, there were only 5 parameters to estimate: Se_{VA}, Sp_{VA}, Se_{SB}, pi1, pi2
- Model run 2: Changing the most likely values to a value still plausible, and increasing slightly the uncertainty of the prior values for the soil sampling bioassay
- Model run 3: Changing the most likely values to a value still plausible, and increasing slightly the uncertainty of the prior values for prevalence

Sensitivity analyses were conducted using one chain and software-generated initial values.

Table 4-2. Changes in prior distributions used for the 3 different models run for the sensitivity analysis (min = minimum, ML = most likely, max = maximum).

Model run	Original priors	Change in prior assumption	Corresponding change in prior distribution	Plot of change in prior distribution (Se/pi1 in green, Sp/pi2 in purple, sensitivity analysis in plain line, original model in dashes)
1	Sp_{SB}: Min 86%, ML 92%	Sp_{SB}=100%	Sp_{SC}=1	
2	Se_{SB}: Min 65%, ML 73% Sp_{SB}: Min 86%, ML 92%	Se_{SB}: min=35%; ML=48%; Sp_{SB}: min=56%; ML=67%;	Se_{SB} ~ beta(1.90, 1.97) Sp_{SB} ~ beta(1.52, 1.25)	
3	pi1: max 80%, ML 50% pi2: max 5%, ML 4%	pi1: max = 60%, ML=40% pi2: max = 30%, ML=15%	pi1~ beta(1.58, 1.86) pi2~ beta(1.28, 2.58)	

4.5 Results and discussion

Ngā hua me te matapaki

4.5.1 Observed test results

The cross-classified test results for the two areas are presented in Table 4-3.

Table 4-3. Number of trees testing positive or negative for *P. agathidicida* by visual assessment (cases) and by soil baiting, culture and morphological identification (*P. agathidicida* detected vs not detected), stratified by population

		P. agathidicida detected (SB positive)	*P. agathidicida* not detected (SB negative)
High prevalence areas (n=189)	Cases (VA positive)	22	26
	Non-cases (VA negative)	35	106
Low prevalence areas (n=572)	Cases (VA positive)	8	73
	Non-cases (VA negative)	11	480

The apparent prevalence, defined as the proportion of tested trees that return a positive test result, of *P. agathidicida* measured by visual assessment were 25.4% in the high prevalence area and 14.1% in the low prevalence area (Figure 4-2 A). The apparent prevalence measured by soil baiting, culturing, and morphological identification was 30.2% in the high prevalence area and 3.3% in the low prevalence area (Table 4-3; Figure 4-2 B).

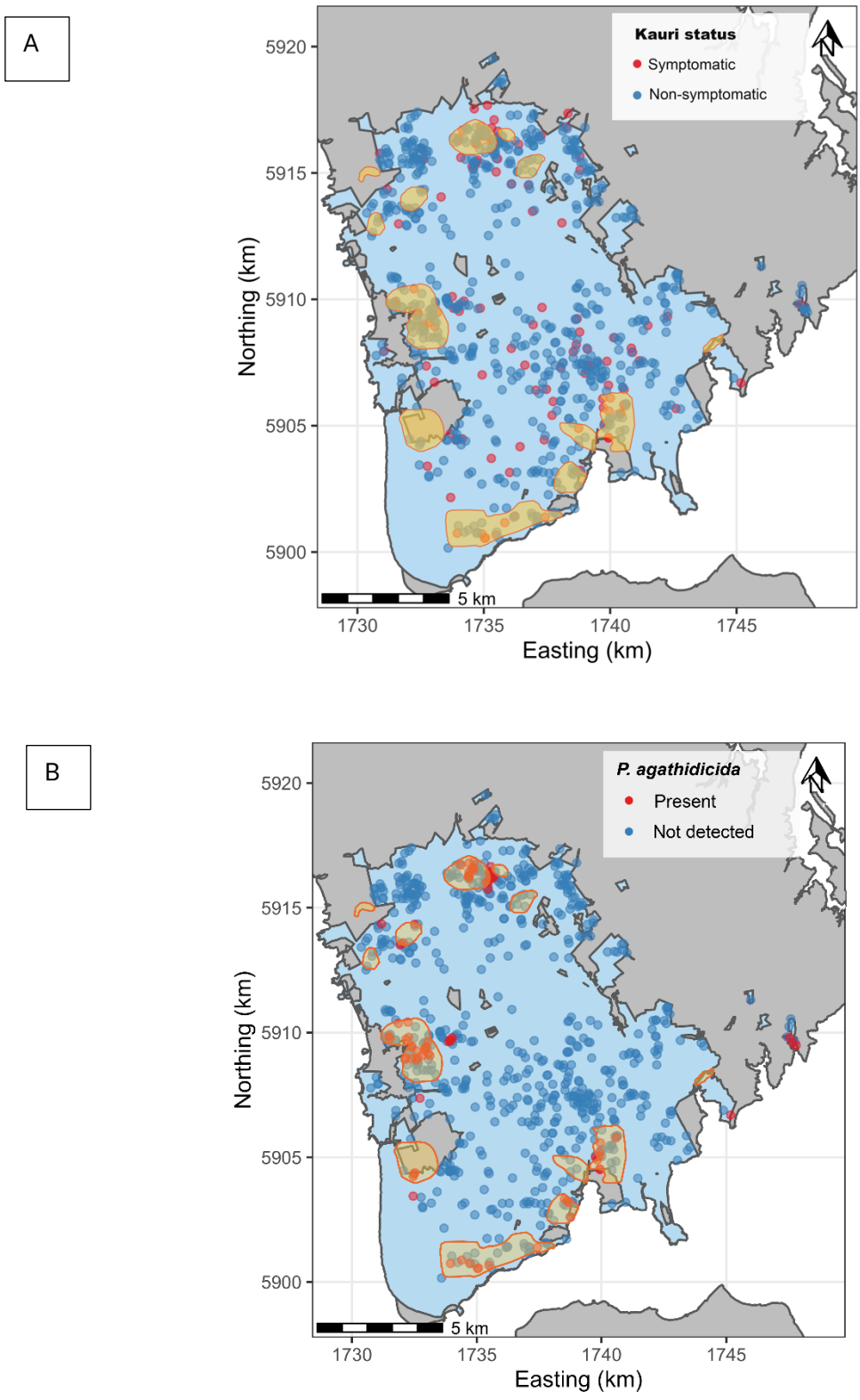


Figure 4-2. Point maps of the 2021 Waitākere Ranges survey showing the prior expected high prevalence areas (yellow-coloured polygons) and A) where *P. agathidicida* was predicted based on the visual assessment test and B) where *P. agathidicida* was detected based on the soil sampling bioassay.

4.5.2 BLCA results

The summary statistics of the posterior distributions for the six parameters are summarised in Table 4-4 and detailed in the following subsections.

Table 4-4. Summary statistics and Monte Carlo error for the six diagnostic test performance and prevalence parameters estimated using Bayesian latent class analysis.

	2.5 percentile	Median	97.5 percentile	Mean	Standard deviation	Monte Carlo error
Se VA	0.2977	0.4096	0.5333	0.411077	0.060471	0.000244
Sp VA	0.8395	0.8699	0.8981	0.869671	0.014984	0.000829
Se SB	0.426	0.6321	0.8809	0.637753	0.116273	0.000059
Sp SB	0.968	0.9872	0.9982	0.986171	0.007943	0.000035
Prevalence (high)	0.3145	0.4641	0.6745	0.471908	0.092095	0.000655
Prevalence (low)	0.01567	0.03804	0.07118	0.039414	0.014253	0.000074

The Monte Carlo error represents the random error that arise because the model takes random draws from probability distributions. In this model they are small, representing less than 1% of the standard deviation of all parameters except the VA specificity for which it is 5.5%. Overall, this means that the summary statistics presented in Table 4.4 are reliable.

4.5.2.1 Visual assessment performance evaluation

The estimated sensitivity for visual assessment was 41.0% (95% PI 29.8-53.3) (Figure 4-3A), which means that less than half of the trees with *P. agathidicida* in the root zone will be recorded positive by visual assessment.

The estimated specificity for visual assessment was 87.0% (95% PI 84.0-89.8) (Figure 4-3B), which means that 13% of trees without *P. agathidicida* in the root zone will be recorded positive by visual assessment.

To help with the interpretation of prevalence studies conducted using visual assessment as described above, the relationship between apparent prevalence (the proportion of trees positive by visual assessment, in other words, the proportion classified as symptomatic trees) and true prevalence (the proportion of truly infected trees, defined here by having *P. agathidicida* in the root zone of the tree) is presented in Figure 4-4. It can be calculated as follows:

TP = (AP+Sp-1)/(Se+Sp-1)

With TP the true prevalence and AP the apparent prevalence (Dohoo et al., 2009). For example, if 30% of trees are positive by visual assessment, the true prevalence of *P. agathidicida* in the soil is 60.7%.

This should, however, be interpreted with caution, as the presence of *P. agathidicida* is spatially clustered, and an estimation of the true prevalence in areas free of *P. agathidicida* would be erroneous. This relationship only applies when apparent prevalence lies between 13% and 41% (Dohoo et al, 2009, p103). If a value outside of these boundaries is observed, it is likely that the sampled trees come from kauri populations that differ from the current study, and the estimated values of sensitivity and specificity don't apply. This is likely to occur frequently with the visual assessment, as in areas (or time points) where dieback is present for other reasons the specificity of visual assessment to detect *P. agathidicida* will decrease.

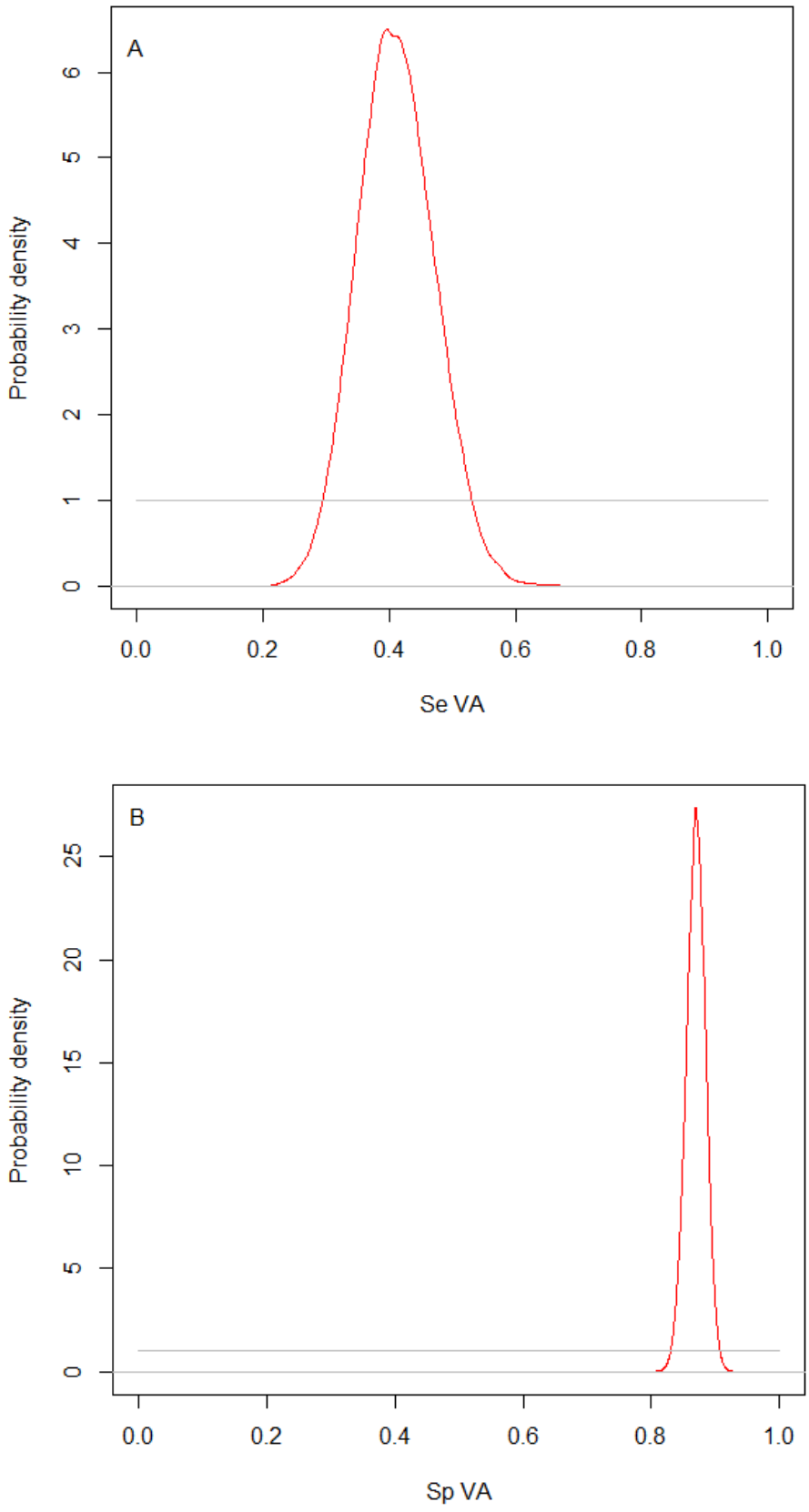


Figure 4-3. Prior (grey) and posterior (red) distributions of the sensitivity (A) and specificity (B) of the visual assessment test for *P. agathidicida*.

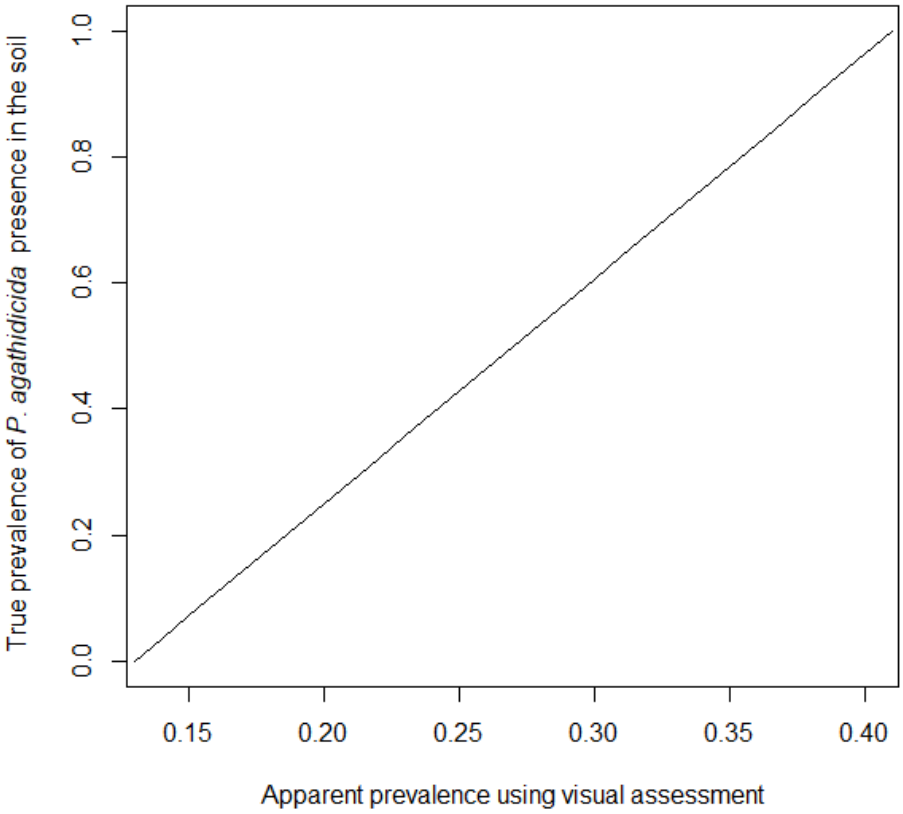


Figure 4-4. Relationship between the apparent prevalence of *P. agathidicida* using visual assessment of disease symptoms, and the true prevalence of *P. agathidicida*.

4.5.2.2 Soil sampling bioassay performance evaluation

The estimated sensitivity for the soil sampling bioassay was 63.2% (95% CI 42.6-88.1) (Figure 4-5A), which means that 63 out of 100 trees with *P. agathidicida* in the root zone will be recorded positive by soil bioassay. This was lower than the value obtained during experts' elicitation.

The estimated specificity for the soil sampling bioassay was 98.7% (95% CI 96.8-99.8) (Figure 4-5B), which was higher than the value obtained during experts' elicitation.

To help with the interpretation of prevalence studies conducted using the soil sampling bioassay as described above, the relationship between apparent prevalence (the proportion of trees returning a positive test result) and true prevalence (the proportion of truly infected, defined here by having *P. agathidicida* in the root zone of the tree) is presented in Figure 4-6. This relationship only applies when apparent prevalence lies between 1.3% (or 0 if we assume a perfect specificity) and 63.2%. In other words, because of the imperfect sensitivity, if the true prevalence is 100%, the apparent prevalence would be 63.2% at the maximum. If a study estimates a prevalence above this number, then the values of sensitivity and specificity calculated here do not apply.

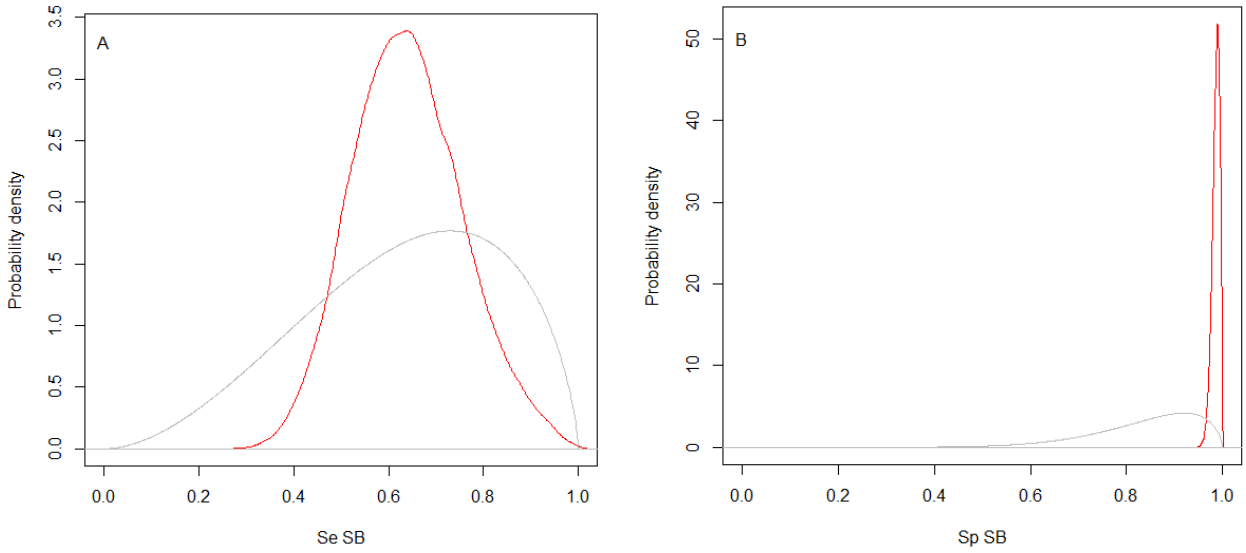


Figure 4-5. Prior (grey) and posterior (red) distributions of the sensitivity (A) and specificity (B) of the soil sampling bioassay test for *P. agathidicida*.

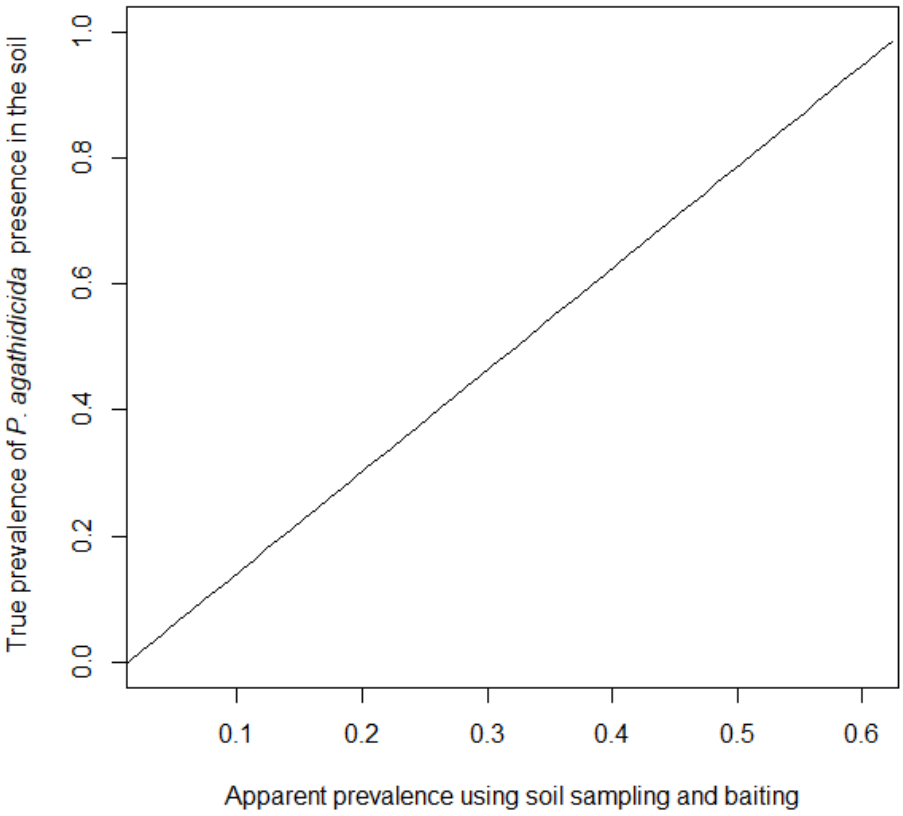


Figure 4-6. Relationship between the apparent prevalence using the soil sampling bioassay, and the calculated true prevalence of *P. agathidicida*

4.5.2.3 Prevalence

The apparent prevalence of *P. agathidicida* measured by visual assessment was 25.4% in the high prevalence area and 14.1% in the low prevalence area (Figure 4-2). The apparent prevalence measured by soil baiting, culturing, and morphological identification was 30.2% in the high prevalence area and 3.3% in the low prevalence area (Table 4-3).

In contrast, the true prevalence estimate, based on the "latent" infection status of the model, for the high prevalence area was 46.4% (95% CI 31.5-67.5, Figure 4-7A), and for the low prevalence area was 3.8% (95% CI 1.16-7.1%, Figure 4-7B). In other words, an estimated 46.4% of the trees in the high prevalence area truly have *P. agathidicida* in their soil. Interestingly, the posterior distribution for the low prevalence area was very similar to the distribution designed from the aerial surveyor's opinion, suggesting that aerial assessment may be an accurate test.

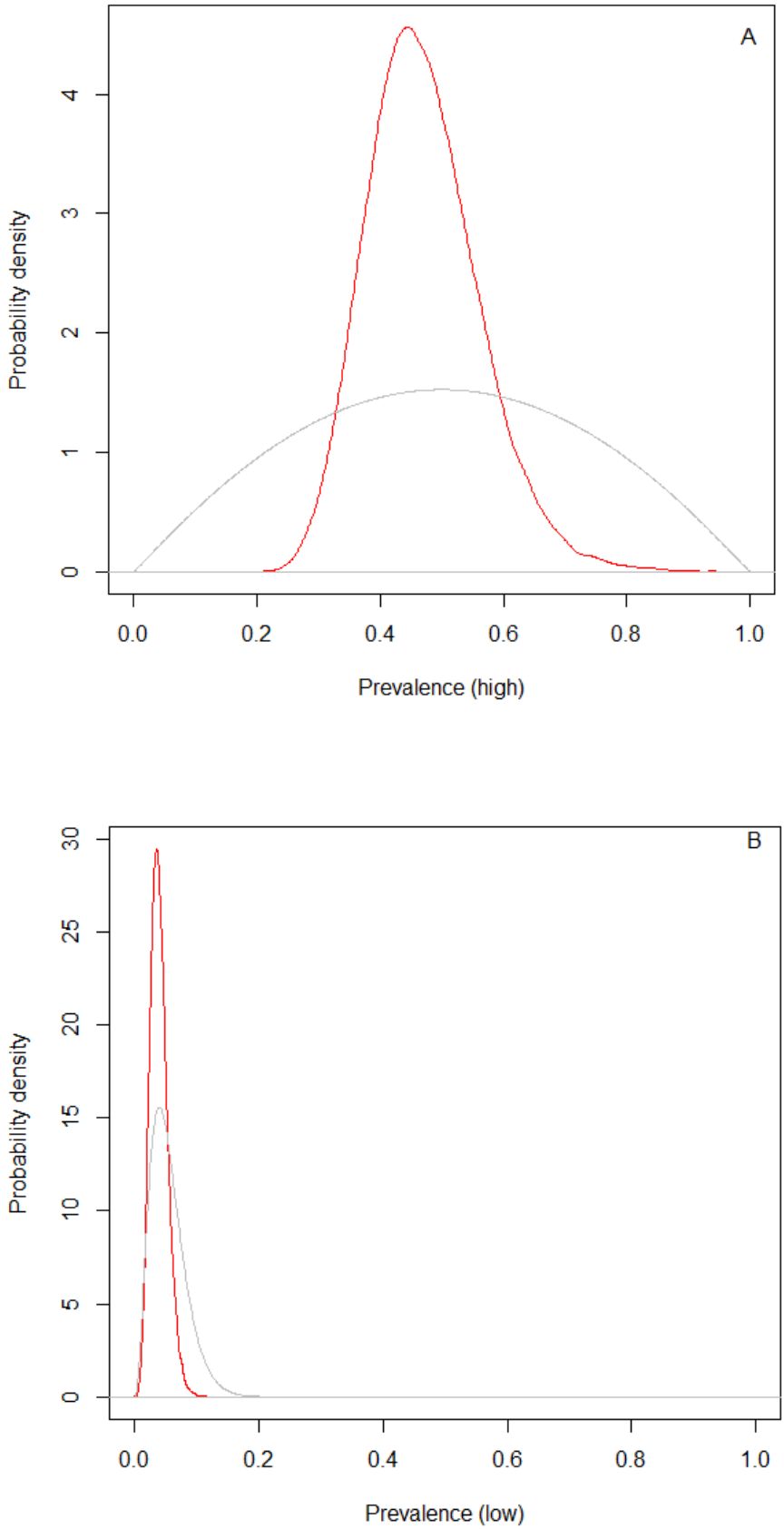
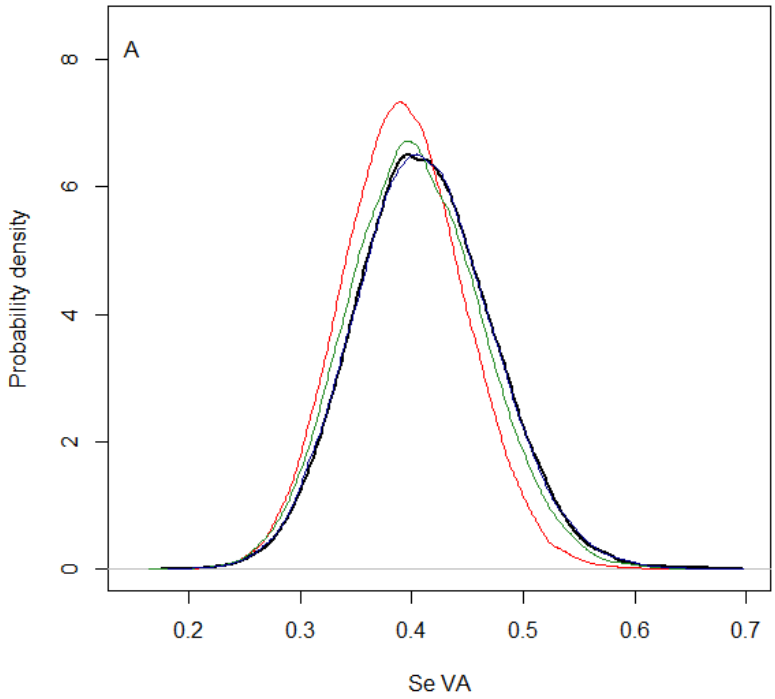
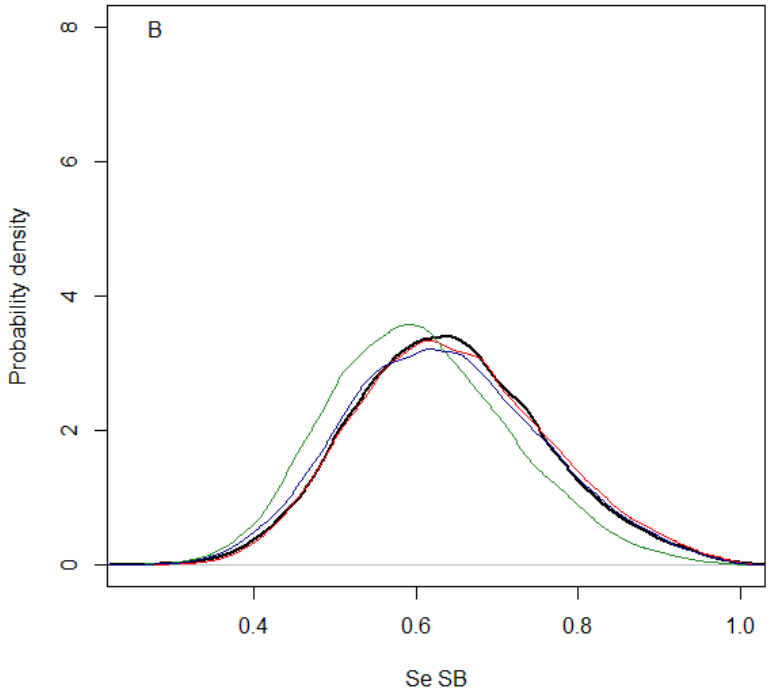


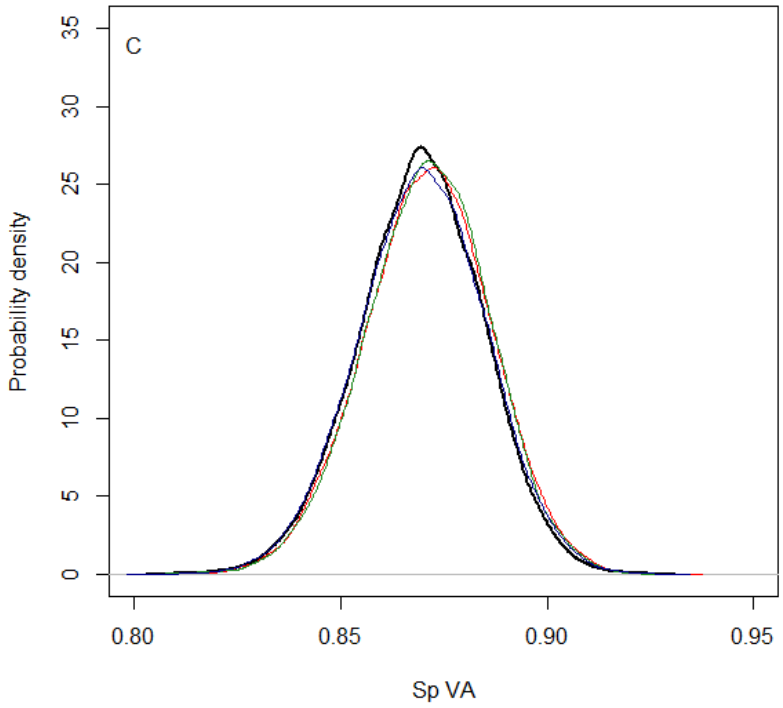
Figure 4-7. Prior (grey) and posterior (red) distributions of the true prevalence of *P. agathidicida* in the high prevalence area (A) and low prevalence area (B).

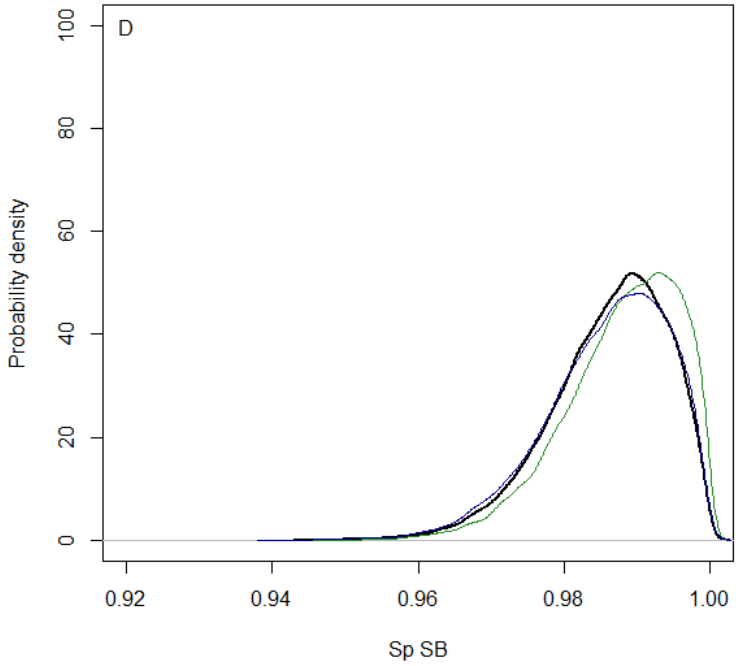
4.5.2.4 Sensitivity analysis

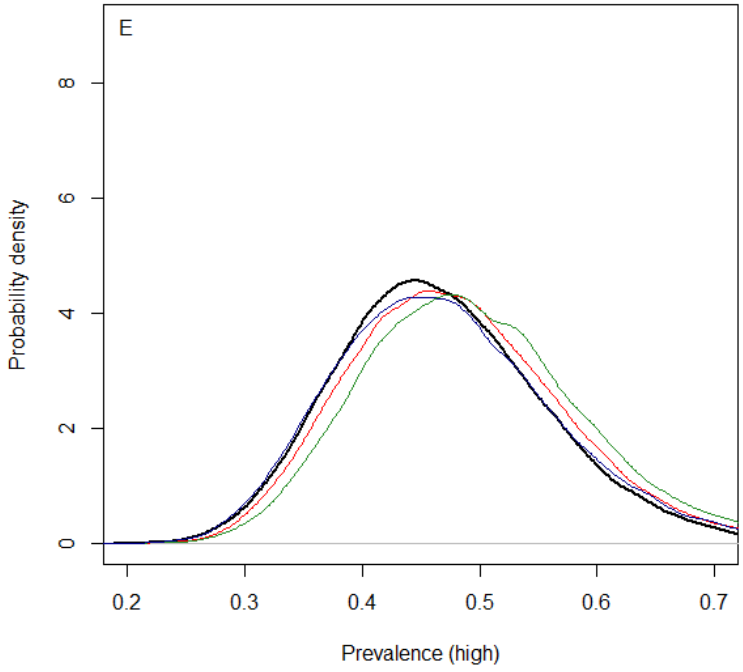
The results of the sensitivity analysis are presented in Figure 4-7. The effect of a change in priors on all the parameters were very small. Assuming a perfect Sp for the soil bioassay (model results shown by a red line on all panels of Figure 4-8) slightly changed the sensitivity of the VA and the prevalence (low), and the soil bioassay sensitivity remained the same, but the precision decreased (63.8%, 95% PI 43.3-89.1). Changing the soil bioassay priors (model results shown by a green line on all panels of Figure 4-8) slightly affected the soil bioassay sensitivity and specificity. A change in the priors for prevalence (in dark blue on Figure 4-8) did not seem to affect any of the parameters.




B
Probability density
0
2
4
6
8
0.4
0.6
0.8
1.0
Se SB


C
Probability density
0
5
10
15
20
25
30
35
0.80
0.85
0.90
0.95
Sp VA


D
Probability density
0
20
40
60
80
100
0.92
0.94
0.96
0.98
1.00
Sp SB


E
Probability density
0
2
4
6
8
0.2
0.3
0.4
0.5
0.6
0.7
Prevalence (high)

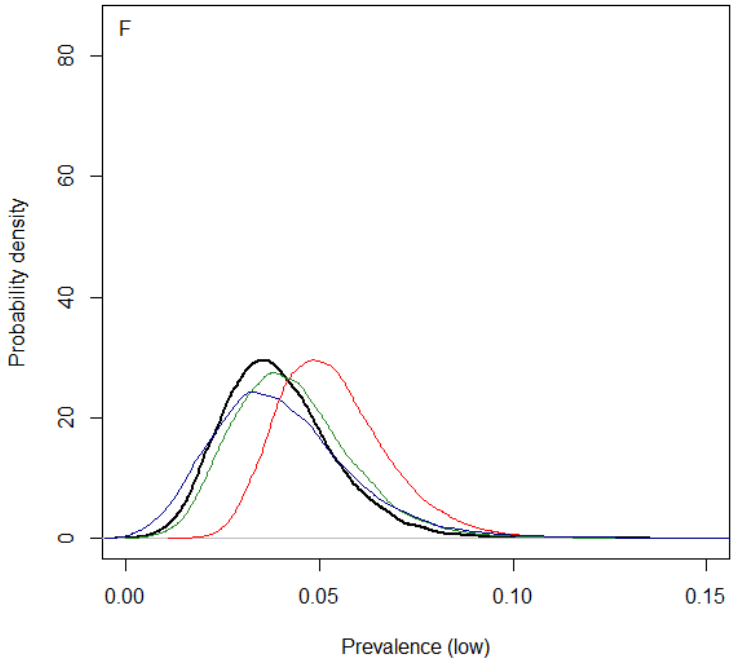


Figure 4-8. Posterior distributions for the sensitivity analysis of the visual assessment sensitivity (A), specificity (B), soil sampling bioassay sensitivity (C), specificity (D), true prevalence in the high prevalence area (E) and low prevalence area (F). The black line was the posterior distribution for the main result using the original priors, the red line for model 1, the forest green line for model 2, the dark blue line for model 3. See Table 4-2 for details on the change in priors for the different sensitivity analysis models.

4.5.3 Limitations

The following methodological limitations were identified and should be considered when using the results of this study.

4.5.3.1 Sampling protocol for soil sampling

It should be highlighted that the results are estimates from the whole procedure from soil sampling to baiting, culture and morphological identification, and not just the laboratory procedure. The methods used to conduct the standard morphological test do not have a standardised measurement of soil for baiting and uses ‘about half a zip-lock sandwich bag of soil’ for baiting. In addition, soil collection in the field used a minimum weight which had high variability due to soil moisture differences on different days (after rain vs long fine periods) and soil composition. It is likely that the test sensitivity depends on the quality of the sample, for example the quantity of fine roots, the composition of the soil, the experience of the person conducting the sampling, storage conditions of the soil along the process, and the volume of soil used in baiting.

4.5.3.2 Assumption of independence

One of the assumptions of the model was that the two tests, visual assessment and the soil sampling bioassay, were conditionally independent. This means that it was assumed that for a kauri tree with *P. agathidicida* in the soil, the knowledge of the visual assessment result would not affect the probability of the soil culture to be positive, and vice-versa; similarly for a tree free of *P. agathidicida*. This assumption is very likely satisfied.

4.5.3.3 Assumption of sensitivity and specificity constant across the study areas

Another important assumption of the model was that the diagnostic sensitivity and specificity of both tests are constant across the different areas and trees sampled. Spatial and temporal variability of the pathogen presence around the tree and soil conditions could affect the sensitivity of the soil sampling and baiting. While there is possibly variability in the samples due to the fact that 16 persons collected soil samples, this was mitigated by specific training and the fact that a large number of trees were sampled. The sensitivity and specificity of visual assessment are also likely affected by other factors such as visibility of the canopy from the ground, experience in using the canopy score scale, in identifying the lesions and in attributing the symptoms to kauri dieback rather than other causes of symptoms around the tree.

4.5.3.4 Sample size

The sample size used in the present study utilised samples from a planned randomised cross-sectional prevalence study (n=189 in the high prevalence area, n=572 in the low prevalence area) and therefore was lower than the recommended sample size in (Vallee et al., 2019). Vallee et al. (2019) recommended at least 800 trees, ideally 1200 across both sites for a specific diagnostic test evaluation study. In addition, only a quarter of the trees in the sample come from the high prevalence area, and the rest come from areas with an overall prevalence estimated at around 4%. This means that the total number of truly infected trees in the sample is likely low, which would have contributed to the higher credible intervals around the sensitivity estimates.

If refining the test sensitivity and specificity estimates was seen as a priority, then a further study could use the current estimates as priors. Indeed, one of the strengths of the Bayesian approach used here is that it can utilise new information to continually refine and improve parameter estimates.

4.5.3.5 Prior distributions

The tests evaluated here differ slightly from the tests for which the prior distributions were established though experts' elicitation (see Vallee et al. (2019)). Because of this, flat priors were used for VA. The SB priors were obtained for soil culture following an 8-point soil sampling protocol, not four points. In addition, the priors for prevalence, were not obtained via formal elicitation, but rather based on the opinion of a single expert, potentially making them more prone to bias. In this study the expert opinion for prevalence was however informed with results from previous studies in the area and was assumed to be reliable. Additionally, the sensitivity analysis showed that a misspecification of the priors for prevalence is unlikely to have affected the results.

4.5.3.6 Implication for interpretation of the tests in series in previous studies

In this study trees were tested using both test methods based on random selection from a sample frame and regardless of disease status. However, previous passive surveillance work has first identified ill-thrift trees during aerial surveillance or as part of a ground survey and then used the two tests sequentially. The trees were first assessed visually and only those with symptoms consistent with kauri dieback were tested by the soil sampling bioassay. Hence, to be considered positive, a tree had to test positive for both tests. Interpreting test results sequentially results in a decrease in diagnostic sensitivity and an increase in diagnostic specificity. This means that a higher proportion of trees with presence of *P. agathidicida* would have been missed by using the two tests in series than by using only one test, notably those that had no symptoms, the "non-symptomatic", but also because of the relatively low sensitivity of the soil sampling bioassay. As an example, in the dataset used for this analysis, in the high *P. agathidicida* prevalence area (where an estimated 189*45.7% = 88 truly infected trees were sampled), *P. agathidicida* was detected using the soil bioassay for more non-symptomatic trees (n=35) than symptomatic trees (n=22); if we had used the two tests in series, we would have classified 35 infected trees as healthy at the visual assessment stage, and thus not tested them for *P. agathidicida*.

If we maintain the original assumption of conditional independence (see Methods section), the sensitivity and specificity of the whole historic sequential testing procedure can be calculated (see Dohoo et al, 2009, p.111). Using the median values obtained from the model with the expert-elicited priors, we obtain the following values:

Se = 25.9%, which means that for 100 trees with a presence of *P. agathidicida*, only 26 would be detected using the sequential procedure

Sp = 99.8%, which means that for 100 trees without *P. agathidicida*, all would almost always test negative using the sequential procedure; in other words, the sequential testing is not expected to have produced any false positives

The previously stated limitations will also apply to these Se and Sp estimates, including the large uncertainty around sensitivity estimates.

These values could in theory be used to calculate the true prevalence of *P. agathidicida* using the apparent prevalence (i.e., the proportion of trees assessed that were positive using the sequential testing; this includes all trees that were visually assessed as not having kauri dieback signs). Some of the historic surveys do not provide the total number of trees visually assessed. If this proportion can be calculated, for example using an estimation of the number of kauri trees in the area, the following formula could be used:

$P = (AP + Sp - 1) / (Se + Sp - 1)$

With P the true prevalence and AP the apparent prevalence. Assuming that Sp = 1, this simplifies as

P = AP/0.259, which is roughly equivalent to multiplying the observed proportion or number of trees by 3.9.

4.5.3.7 Implications for sample size calculations for freedom of P. agathidicida

Using the sensitivity estimate (63.8%) assuming perfect specificity (100%) for the soil sampling and bioassay obtained from Model 1 in the sensitivity analysis, the number of trees to be sampled from an area to demonstrate freedom of *P. agathidicida* can be easily calculated, for example using a calculator such as https://epitools.ausvet.com.au/freedomss.

For example, in an area with 10,000 kauri trees, you would need to test 47 trees by the soil sampling bioassay to detect *P. agathidicida* at a prevalence of 10%, 94 at a prevalence of 5%, or 463 at a prevalence of 1%. If all trees return a negative result, we would be 95% confident that if the pathogen is present, it would be below this "design" prevalence. Note that the design of such a study would need to focus on areas that are small enough to assume a homogenous distribution of truly infected trees. The sample size will change depending on the level of confidence desired (the "required population sensitivity") and the population size of kauri trees in the area.

4.6 Conclusions and recommendations

Te whakatau me ngā tūtohunga

This study used data from a cross-sectional study to estimate diagnostic sensitivity and diagnostic specificity of visual assessment and soil sampling bioassay to detect the presence of *P. agathidicida* in the soil around kauri trees, using Bayesian latent class analysis. The study area was divided into presumed high and low prevalence areas.

For visual assessment, the estimated sensitivity was 41.0% (95% PI 29.8-53.3) and the estimated specificity 87.0% (95% PI 84.0-89.8). For the soil sampling bioassay, the estimated sensitivity was 63.2% (95% PI 42.6-88.1) and the estimated specificity 98.7% (95% PI 96.8-99.8). If we assumed a perfect specificity, the sensitivity for the soil sampling bioassay was 63.8% (95% PI 43.3-89.1). These values can be used to calculate the true prevalence of *P. agathidicida* in past and previous studies that used visual assessment or the soil sampling bioassay using the same procedures, or a sequential use of both tests. When both tests were used in series, it was estimated that the true prevalence was underestimated by a factor 3.9 in historical studies. These values, especially the diagnostic sensitivity of the soil sampling bioassay assuming a perfect specificity, can be used to calculate the required sample size for a proof-of-freedom survey.

The pre-defined high prevalence area had an estimated true prevalence of *P. agathidicida* of 46.4%, and the remaining low prevalence area had an estimated true prevalence of 3.8%.

The values obtained in this study are valid only for tests conducted using the same test methodology, i.e., with visual assessment following the exact same procedures, by skilled operators, or the soil sampling bioassay using the exact same soil collection methodology and laboratory procedures. An assessment of operator agreement would also be useful in deciding if an overall value is sufficient or if operator or laboratory specific values are needed. These results can be used as informed priors for future refinement of the sensitivity and specificity parameters. It is also recommended to interpret test results for prevalence studies on limited areas where the distribution of pathogen presence can be considered homogenous.

It is recommended that current and future tests' accuracy are also evaluated using Bayesian latent class analysis, which allows demonstrating a higher sensitivity of new tests, that the gold standard method does not allow.

Find out more: kauri@aucklandcouncil.govt.nz
or visit knowledgeauckland.org.nz and
aucklandcouncil.govt.nz

Auckland Council
Te Kaunihera o Tāmaki Makaurau