Report Cover Page

ACERA Project
1004

Title
Post-border surveillance techniques: review, synthesis, and deployment

Author(s) / Address(es)

Dr Susie Hester, School of Business, Economics and Public Policy, University of New England, Armidale, NSW

Dr Evan Sergeant, AusVet Animal Health Services, Orange, NSW

Dr Karen Herbert, Biosecurity Victoria - Department of Primary Industries, Albury, NSW

Dr Andrew Robinson, Department of Mathematics and Statistics, University of Melbourne, VIC

Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)
Final Report 4 (Stage 5)

Summary

Post-border surveillance is used to give evidence that a pest or disease is absent from a country, region, or defined area, thus enabling access to particular export markets; detect new pests and diseases early enough to allow for cost-effective management; establish the boundaries of a known pest or disease; and monitor existing containment or eradication programmes.

A variety of tools exists to aid biosecurity managers plan, implement, and evaluate post-border surveillance activities. These were reviewed in Stages 1 and 2 of this project. Many of the tools and methods discussed in the review are, however, not easily applied by those involved in post-border surveillance due to both the complexity of the tools and time constraints on surveillance staff.

Previous milestone reports outlined two case studies that illustrate the application of two of these tools (Stage 3), and described their implementation in ways that would make them accessible to operational staff in Australian government agencies (Stage 4). The purpose of this report is to describe field tests of the tools, recommendations for modifications and developments to suit operational conditions, and (for case study ii) the test version of the software (Project Stage 5).

In summary, the two case studies explain:

i. the use of *EpiTools* (a set of web-based tools) to create a survey strategy for demonstration of freedom from citrus canker in the Northern Territory; and
ii. the use of an Excel-based eradograph-monitoring tool, to show progress towards regional extirpation of orange hawkweed in the Australian Alps.

ACERA Use only	Received By:	Date:
	ACERA / AMSI SAC Approval:	Date:
	DAFF Endorsement: ( ) Yes ( ) No	Date:

Post-border surveillance techniques: review, synthesis, and deployment

ACERA Project No. 1004

Susan Hester, University of New England

Final Report 4 (Stage 5)

15 December, 2011

une
University of
New England

Acknowledgements

Table of contents

List of Tables

List of Tables

List of Figures

1. Executive Summary

A variety of tools exists to aid biosecurity managers plan, implement, and evaluate post-border surveillance activities. These tools were reviewed in Stage 1 and 2 of this project and range from rules of thumb and formulae to user-friendly interfaces for simulation models. Many of the tools and methods discussed in the review are, however, not easily applied by those involved in post-border surveillance due to both the complexity of tools and the time constraints on surveillance staff who would be required to understand and apply them.

Previous milestone reports outlined two case studies that illustrate the application of two of these tools (Stage 3) and described their implementation in ways that would make them accessible to operational staff in Australian government agencies (Stage 4). The current report describes field tests of the tools using case studies, it contains recommendations for modifications and developments to suit operational conditions, and a description of the Excel-based tool (Case Study 2) (Project Stage 5).

The first case study explains the use of *EpiTools*, a pre-existing set of web-based tools, to create a survey strategy for demonstration of freedom from citrus canker in the Northern Territory. EpiTools can be used to design surveys that meet market access requirements. This set of tools has been well applied in the animal sector, but there has been little or no uptake of it in the plant sector despite applicability of the tools to plant-health surveillance problems.

The second case study explains the development and use of an Excel-based eradication-monitoring tool, incorporating an 'eradograph' to show progress towards regional extirpation of orange hawkweed in the Australian Alps. This tool allows biosecurity managers to improve the monitoring of the effect of weed management activities, and evaluate progress in an eradication programme as a basis for making sound decisions on the future delivery of such programmes.

This study demonstrates how the tools would be used in situations typical of those faced by plant-health managers. Both tools are ready to be applied operationally and can improve the capability of agencies tasked with undertaking surveillance, but with limited expertise and resources, to deliver sound and defensible surveillance biosecurity outcomes for Australia.

The use of EpiTools is recommended:

1. Where a structured survey is required to prove freedom in a plant-health context, to design surveys that will generate a required level of confidence (e.g. 95%) of detecting a disease/pest at or above a specified prevalence (e.g. 1%);

2. Where the budget for a structured survey is limited, to find the least-cost sample size that would be required in order to generate a particular level of confidence (e.g. 95%) of detecting a disease/pest at or above a specified prevalence (e.g. 1%).

In either case, survey designs could then be reviewed by a statistician if required.

The use of the eradication monitoring tool, incorporating the eradograph, is recommended:

1. Where an objective ongoing measure of the progress of a weed eradication programme is needed to assist decision making on future delivery of the programme.

2. Introduction

Surveillance for a range of exotic pests and diseases is routinely undertaken by biosecurity managers across Australia, for reasons of market access, early detection, delimitation, and monitoring. Information derived from these surveillance activities is used in making decisions about future management incursions.
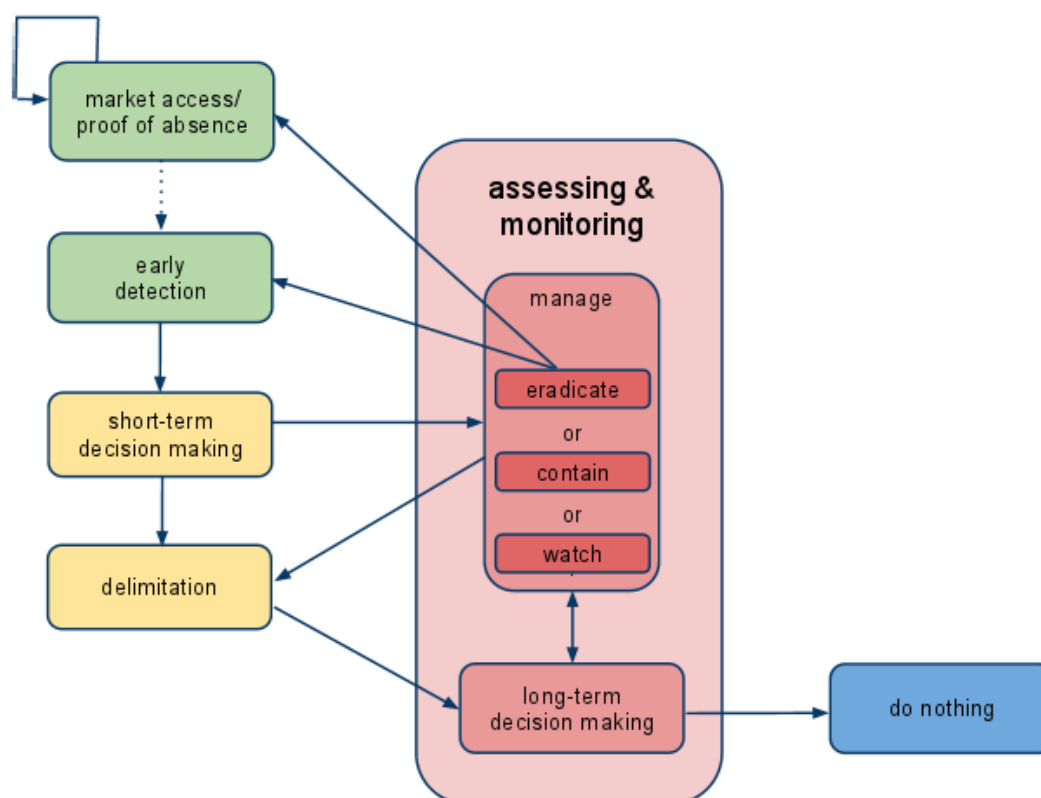
Figure 1. Conceptual diagram showing the phases of surveillance and infestation management

The post-border surveillance schema, illustrated in Figure 1 (from Hester *et al.* 2010), explains how these post-border surveillance activities fit together. Detections of a pest or disease that result from surveys undertaken when a pest of disease is thought to be absent (market access and early detection) lead to some short-term decision making to determine the appropriate initial response. In some cases, a protocol may have been agreed upon prior to detection (e.g. PHA 2006; 2007; AHA 2008a) and management can proceed immediately. Alternatively or simultaneously, delimitation may be required to understand the full spatial extent of the incursion. Knowledge of the current and potential extent of the incursion would allow estimation of the level and value of the damages that the incursion might cause and the resources required for particular management strategies. These strategies might be to eradicate, to contain, or simply to watch the incursion with little interference. Over time, further delimitation surveys may be required in the evaluation of the management

programmes and, depending on the outcome, management strategies might change (see, for example, Rout *et al.* 2010, and Moore *et al.* 2011). The additional management option (viz. do nothing) might be appropriate if further active management cannot be justified.

The tools and methods that can assist with decision making in the context of the varied aspects of post-border surveillance were summarised in Hester *et al.* (2010). Often, the available methods and tools are not readily applied because of a lack of skills and/or time constraints on surveillance staff who would be required to understand and apply them. Discussions with biosecurity managers identified that survey design and evaluation of eradication programmes were two areas where usable tools could be developed that may lead to significant benefits. In consultation with a group of biosecurity managers[1] the following case studies were developed and are discussed in this report along with the tools that are field tested in each case:

i.   Application of EpiTools (Sergeant 2009) to the design of a statistically sound citrus canker surveillance strategy for the Northern Territory; and

ii. A tool that allows progress towards eradication to be quantified (Burgman *et al.*, submitted), and applied to the extirpation of orange hawkweed in the Australian Alps, Victoria.

The over-arching objective of this multi-stage project is to identify and apply tools whose application will result in efficient allocation of resources among competing biosecurity risks to provide maximum public benefit. To achieve this over-arching objective, the project has been divided into six stages:

- Stage 1: Review and synthesise ACERA research;

- Stage 2: Review and synthesise national and international research;

- Stage 3: Develop scenarios, case studies, and examples that illustrate the application of tools in circumstances relevant to their deployment in operational conditions in Australia, with end-user involvement;

---

[1] On 26 and 27 October 2009, participants from ACERA (Susie Hester, Andrew Robinson, Paul Pheloung, Mark Burgman) met with colleagues from the NT Department of Resources (NT DoR) (Andrew Tomkins, Sue Fitzpatrick, Helen Cribb, San Kham Hornby, Graham Schultz, Jim Swan) to discuss surveillance needs. On 10 and 11 August 2010, Susie Hester discussed surveillance needs with biosecurity managers from Qld, NT, and Northern Australia Quarantine Service (NAQS) at an Australian Biosecurity Intelligence Network (ABIN) workshop on biosecurity in northern Australia.

The second case study was suggested by Fran Hausmann and developed in consultation with Karen Herbert, both of Biosecurity Victoria.

- Stage 4: Develop and test simple software and spreadsheet applications that will facilitate the use of these tools in standard operating conditions in federal and state agencies;

- Stage 5: Guide development of these tools by testing them iteratively in field conditions, and modifying the tools as required to suit a range of operational conditions, with end-user involvement; and

- Stage 6: Develop guidelines and training materials and provide training opportunities for these tools (coordinating with the ACERA project for training in risk analysis tools).

In this document we present the results from Stage 5 (field tests of the tools, recommendations for modifications and developments to suit operational conditions, and the test version of the software).

Application of EpiTools to the design of a citrus canker surveillance strategy is a collaborative effort between the Northern Territory Department of Resources (NTDoR), the Northern Australia Quarantine Service (NAQS), and ACERA. Data on citrus surveillance in the Northern Territory has been provided by NTDoR. This data has been used in the case study to demonstrate how to EpiTools may be used to provide a statistically sound survey strategy. Instructions for using EpiTools in this way should facilitate its further use in determining surveillance strategies for additional pests and diseases of plants.

The second case study—applying the eradograph tool to the extirpation of orange hawkweed in the Victorian Alps—is a collaborative effort between Biosecurity Victoria and ACERA.

3. Case Study 1: Survey-design tool – citrus canker

Susie Hester, Evan Sergeant, and Andrew Robinson

3.1. Background

When surveillance is undertaken to establish or maintain market access, biosecurity managers are required to use science-based evidence to support their claims that a pest or disease is absent from a country or region. This evidence is provided by using population-based surveys, non-random (purposive) surveillance, or general/passive surveillance. Where purposive surveys are undertaken and no pest is found, the results are used to show that there is a particular level of confidence (e.g. 95%) that the pest would have been found even if it were present at a very low prevalence (e.g. 0.05%).

For some pests and diseases, regulations exist for how disease freedom should be demonstrated. Where this is the case, statistical requirements for survey design and guidelines for non-random surveillance are specified and there may be little scope for deviation from these (e.g. scrapie surveillance; AHA 2008b). In the absence of prescribed rules, biosecurity managers are responsible for designing the surveys that are used to demonstrate pest absence. This design process involves determining the number of locations to measure, choosing the locations from which survey information is collected (the sampling plan), and the number of units within each location that will be sampled (sample size). The choice of sampling plan may be influenced by prior information about the locations and by their spatial distribution, and sample size is influenced by the effectiveness of the testing method, the confidence required, and the available budget.

Discussions with biosecurity managers in northern Australia revealed that they seldom have the time, and they lack the statistical skills, to design the surveys that are required to support claims of area freedom.[2] As a result, an alternative strategy of surveying the entire population of known hosts is often chosen, which leads to unnecessary expenditure if surveying only a subset would have been adequate. Biosecurity managers who do not have the time or skills to design appropriate surveys would benefit from a tool that they could use to determine:

- the number of host animals/plants/locations that should be checked to enable a certain level of confidence that if the pest/disease is present, it would be found;

---

[2] On 26 and 27 October 2009, participants from ACERA (Susie Hester, Andrew Robinson, Paul Pheloung, Mark Burgman) met with colleagues from NT DoR (Andrew Tomkins, Sue Fitzpatrick, Helen Cribb, San Kham Hornby, Graham Schultz, Jim Swan) to discuss surveillance needs. On 10 and 11 August 2010 Susie Hester discussed surveillance needs with biosecurity managers from Qld, NT, and Northern Australia Quarantine Service (NAQS) at an Australian Biosecurity Intelligence Network (ABIN) workshop on biosecurity in northern Australia.

- how survey information could be used to robustly estimate the likelihood that the pest/disease is not present; and

- the level of resources needed to meet the survey requirements to ensure market access.

An existing set of web-based tools (EpiTools, Sergeant 2009) has been developed to support survey designs for estimating disease prevalence or demonstrating freedom from diseases in animal herds. While these tools work well for their designed purpose, and should be applicable to citrus canker surveillance, they appear not to be widely used in plant-health surveillance. Furthermore, the tools are intended for use by epidemiologists and other researchers who have a good understanding of statistical terminology and use of statistical concepts. These skills are not universal among biosecurity managers.

In this case study, we will demonstrate the use of several of the statistical functions provided in EpiTools by designing a citrus canker survey strategy for the Northern Territory. Citrus canker, the proposed survey-design tool, data, and the results from applying the tools are now discussed.

3.2. Citrus canker

Citrus canker is a highly contagious disease of citrus trees (grapefruit, limes, lemons, and oranges) caused by the bacteria *Xanthomonas axonopodis* pathovar *citri*. Infected trees suffer from low vigour, and in serious cases, maturity is delayed. Leaf, stem, and fruit blemishing lead to a reduction in the quantity and quality of fruit produced by infected trees (Figures 2A and 2B). Fruit from infected trees is scarred and usually cannot be sold. Further, interstate and export markets only accept citrus fruit that is produced in areas that are free of the disease.
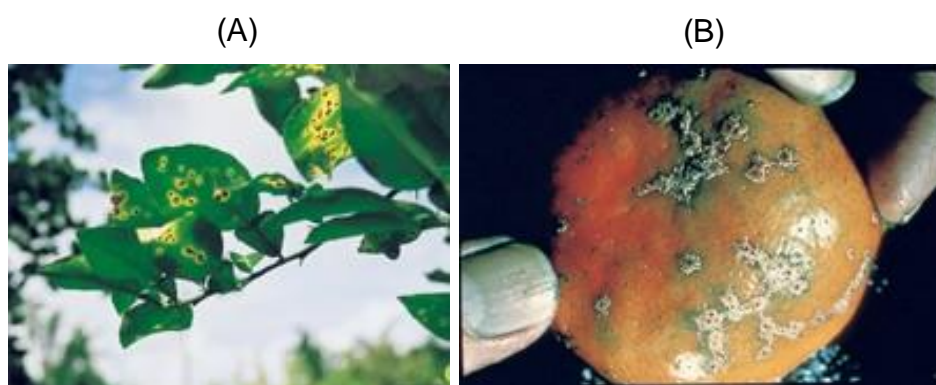
(A)   (B)

Figure 2. Photos of citrus canker present on leaves (A) and fruit (B) of a citrus tree. Photos: http://www.daff.gov.au/aqis/quarantine/naqs/naqs-fact-sheets/citrus-canker

Citrus canker is common in countries to the north of Australia, including Indonesia and Papua New Guinea, so it is a high-priority plant pest for the Australian citrus industry (PHA 2009). As a result, surveillance for citrus canker in both urban and non-urban areas of Northern Australia is undertaken as part of the Northern Australia Quarantine Strategy (NAQS). In addition to this targeted surveillance, the State and Territory governments in Northern Australia conduct a range of additional surveillance activities. Information on citrus canker may also result from general surveillance, where agronomists, consultants, and citrus growers provide information on the health of citrus trees with which they come into contact.

Australia was recently declared free of citrus canker following a four-year programme to eradicate the disease from around the township of Emerald, Queensland, where it had been detected in three commercial citrus orchards between June 2004 and May 2005 (DAFF 2009). A protocol for citrus canker surveillance of production orchards was developed following the outbreak in Emerald. For production areas outside Queensland, the protocol contained details of surveys for detection of citrus canker that would be appropriate to defend the claim of pest-free-area status to international markets. Specifically, the sampling protocol was '...designed to detect a level of 1% or more of host material infected with *Xanthomonas axonopodis* pv. *citri* on a growing site. It provides a 95% confidence in detecting the pathogen at locations where the percentage of infested hosts is at least 1%' (FAO 2002 cited in OCCPO 2004). While this protocol is useful to determine how surveys should be performed in commercial orchards, it ignores citrus trees that are not grown in productive orchards, and so excludes a large number of citrus trees in the Northern Territory. There is, therefore, a need for survey designs that include both commercial and non-commercial citrus trees.

Maintaining area-freedom status for citrus canker in order to provide access to major overseas markets is a high priority for the Australian citrus industry (PHA 2009) whose exports of citrus were valued at $156 million in 2008/09 (ABARE 2009). Surveillance for citrus canker in the Northern Territory is also undertaken in order to comply with area freedom requirements from interstate trading partners.

3.3. EpiTools: key terminology and concepts

EpiTools is a set of web-based tools that may be used to develop structured surveys for use in estimating disease prevalence or demonstrating freedom from diseases. The statistical tools that are provided on the EpiTools website use design-based, frequentist sampling theory, where inferences drawn from the data are derived from the large-sample characteristics of the sample design, rather than any prior ideas of probability density

functions or models for the data. Key statistical terms used in EpiTools are explained in Table 1 and important formulae used by the functions in EpiTools listed in Appendix 1.

EpiTools is located on the AusVet Animal Health Services website, located at http://www.ausvet.com.au/, under the menu item **Tools** (Figure 3A). Once EpiTools is selected from the list of options that are displayed when **Tools** is highlighted, the EpiTools home page appears (Figure 3B).

When translating EpiTools from an animal-surveillance context to citrus canker, it is useful to think of herds or farms as being equivalent to backyards or orchards, and animals as being equivalent to trees.
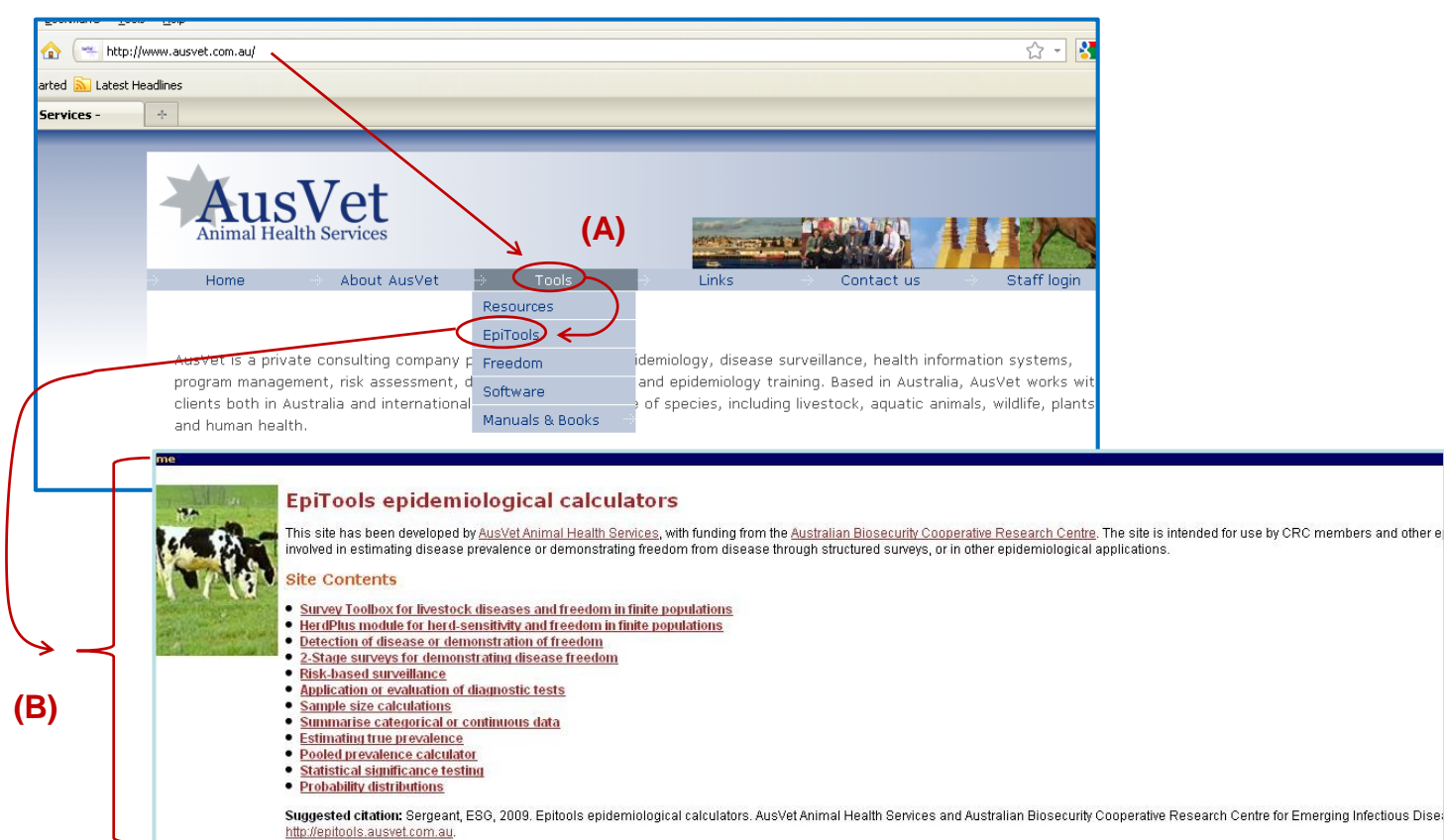

Figure 3. Screen view of how to locate EpiTools on the web: (A) shows the location of EpiTools on the AusVet home page, (B) shows the front page of EpiTools

Table 1. Definitions and values of key concepts used by EpiTools

Parameter, animal-surveillance context (as in EpiTools)	Parameter, plant-surveillance context	Description in citrus canker context	Symbol	Value in citrus canker context (Source)
Test sensitivity	Test sensitivity	The diagnostic sensitivity of a test. This is the probability that an individual diseased tree will be correctly identified as positive by the test. Also called True Positive Rate (of a test). When calculating system sensitivity or number of orchards to sample for two-stage sampling, use location (a population of trees in a defined space), or orchard sensitivity (see below).	Se	0.5 (no information available on this value, so a value of 0.5 assumed – standard practice in this situation)
Herd sensitivity	Location or orchard sensitivity	The probability that an infected orchard/location will give a positive result following a particular testing protocol, given that the disease is present in trees at a prevalence equal to or greater than the design prevalence.	SeH	0.95 (OCCPO 2004)
Design (target) prevalence	Design (target) prevalence	This is a pre-survey hypothetical level of disease that a survey is designed to detect, measured as the proportion of the total number of host trees at a location or in an orchard that have citrus canker (tree level), or the proportion of orchards or properties that have the disease (orchard level). Design prevalence can be applied at the tree or orchard levels or both (see below).	P*	0.01, 0.005 (NTDoR staff, OCCPO 2004)
Herd-level design prevalence	Orchard-level design prevalence	The hypothetical proportion of diseased orchards or properties that a survey is designed to detect (assuming each property is diseased at or above the tree-level design prevalence).	P*	0.01 (OCCPO 2004)
Animal-level design prevalence	Tree-level design prevalence	The hypothetical proportion of diseased trees in a population (either a specific location or property or a broader population of trees) that a survey is designed to detect.	P*	0.01 (NTDoR staff, OCCPO 2004)
System sensitivity	System sensitivity	The overall probability (level of confidence) of detecting disease if it is present in the population at the specified design prevalence(s). May be specified as a target to be achieved or calculated as the actual level achieved by the survey.	SSe	0.95 (NTDoR staff, OCCPO 2004)

3.4. Designing a citrus canker surveillance strategy using EpiTools

To start the process of designing a surveillance strategy it is important to understand which questions the planned survey activities are designed to answer. Typically, area freedom is agreed between trading partners to be practically demonstrated if a sampling strategy will provide a high level of confidence of detecting the pest/disease at a low (but greater than 0) prevalence. In the case of citrus canker, the questions that are being asked by biosecurity managers in the Northern Territory are:

- How should survey locations for citrus canker be selected across the Northern Territory, in order to provide 95% confidence that the disease will be detected where the prevalence across those locations is at or above a specified low level (say, 1%), and in accord with reasonable expectations about the spatial pattern of the infestation?

- How many trees within each location should be sampled to ensure 95% confidence that citrus canker would be detected where the prevalence in trees within the location is at or above a specified low value (say, 1%)?

3.4.1 The citrus canker dataset and its configuration for use in EpiTools

All citrus trees are considered to be potential hosts for citrus canker. In the Northern Territory, citrus trees are located across a wide range of land types and at a range of densities, from a few trees in suburban backyards and isolated trees in remote communities to large numbers of trees in commercial orchards.

Data on citrus surveillance provided by the Northern Territory government for this case study were collected during a Territory-wide 2005-6 survey and consist of the following:

- location of the tree(s) (including the suburb/district, street number, and address, LTO or Sec No);

- the type of property containing the tree(s) (comprising the categories: not recorded, vacant, crown vacant land, residential, commercial, nursery, industrial, or light industrial);

- the number of citrus tree(s) at each location, recorded as either an exact number or simply that there were trees present (without an exact number);

- whether or not the planting is classed as a plantation (an orchard or small planting of trees).
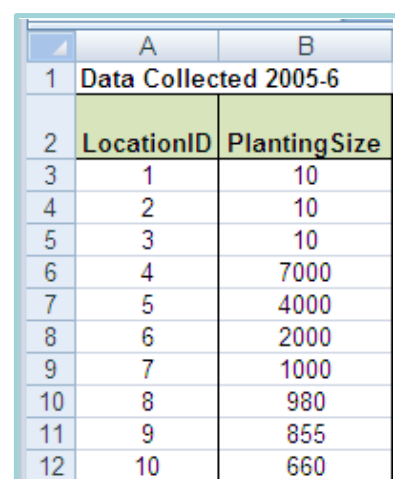
The original dataset, contained in an Excel spreadsheet, consisted of records from 1497 locations, many of which were recorded as having no citrus hosts present. Since the objective of this analysis was to decide which trees should be surveyed, only locations containing hosts were used in the analysis, totalling 408 locations (>23,000 trees).

For three locations where trees were recorded as being present at a location but for which there was no record of the number of trees, a value of 10 was inserted to ensure they were represented in the study. Ten was chosen as a reasonable number of trees for unknown locations given that about 85% of locations had ≤ 10 trees recorded.

It is useful to know that several of the tools used in this analysis require the data to be organised so that a column containing a location identifier and a column containing the number of trees at the location are placed next to each other (Figure 4). Additional columns of data can be included if desired (after the two columns of LocationID and PlantingSize) and rows can be in any order. Detailed outputs will be in the same order as input and will also include any additional columns provided.

In summary, the following steps were taken to organise the dataset for use in EpiTools:

- **locations containing zero host trees were removed from the dataset** (408 locations remained)**;**

- **three locations where host trees were present, but where the exact number of hosts was not specified, were given a value of 10 trees;**

- **a column that gives each location a unique identifier was inserted into the dataset, called *LocationID,* and its cells numbered from 1 to 408** (Figure 4); and

- **a column containing the number of trees at each location was inserted immediately to the right of *LocationID* and named *PlantingSize*** (Figure 4).

Figure 4. A partial view of the two columns that will be used for analysis in the two-stage surveys in EpiTools.

3.4.2 A one-stage survey for citrus canker

In a one-stage survey, a random sample is collected from the whole population of interest, regardless of the fact that some trees are clustered in various locations such as orchards. One-stage sampling is an appropriate survey method for those situations where every

member of the population is known, and can be listed and located. The list of all the population units is called the sampling frame. This survey method involves calculating an appropriate sample size using a standard statistical formula[3] and then selecting individuals for testing from the sampling frame. Here, individuals are selected using simple random sampling, a sampling technique in which every possible *n*-sized combination of members of the population has the same probability of being selected. We outline alternative approaches to selecting the sample in Section 3.6.

The calculation of the appropriate sample size is based on the performance of the test (its sensitivity), a pre-survey estimate of the target proportion of infested individuals to be detected (design prevalence), and the desired system (or population) sensitivity (the overall level of confidence of detecting disease if it is present). Because the survey is being carried out to demonstrate freedom from disease, the design prevalence is expected to be close to zero. If the true prevalence is higher than the design prevalence, then the design will be conservative; that is, a higher number of samples will be prescribed than are needed for the objectives of the study, and *vice versa*.

Once the sample size has been calculated, random sampling is used to determine which trees from the population will actually be surveyed. EpiTools provides several tools that could be used to do this, depending on population size.

To use EpiTools to calculate the sample size for a one-stage survey for citrus canker these steps should be followed:

1. Select **Detection of a disease or demonstration of freedom** from the EpiTools home page (Figure 5A);

2. If the population size is unknown, but it can be assumed that it is large, select **Sample size for demonstration of freedom in a large population** from the list of options that now appear on the screen (Figure 5B). If the population size is known, the option **Sample size for demonstration of freedom in a finite population** should be used instead;

3. Insert values for **test sensitivity** (0.5), **desired herd sensitivity** (0.95), and **design prevalence** (0.01) into the appropriate input box (Figure 5C). Values in parentheses are for citrus canker and are based on the values used for a **block[4]** in the 2004

---

[3] Sample size formulae used in EpiTools are provided in the Appendix for information. Where the population is large relative to sample size, or population size is unknown, the binomial method can be used; for small populations or where population size is known the hypergeometric is preferred

[4] In the 2004 national survey (OCCPO 2004), a block was defined as a contiguous group of host trees managed by one producer—the number of trees in a block must exceed 500 and ideally is about 2000.

national survey. Reducing the design prevalence to 0.001 or even 0.005 might be more realistic, but would result in much higher sample sizes as explained below.

Figure 5. Screen view of steps involved in determining sample size using a one-stage survey: (A) view of the EpiTools home page; (B) view of screen choices available when *Detection of disease and demonstration of freedom* is selected from home page; (C) inputs required for determining *Sample size for demonstration of freedom in a large population*; and (D) results from the chosen parameter values.

Figure 6. Screen view of steps involved in taking a random sample of the data: (A) view of the EpiTools home page; (B) view of screen choices available when *Survey Toolbox for livestock diseases and freedom in finite populations* is selected from home page; (C) inputs required for *Random sampling of animals from a list of owners*; and (D) the area where HerdID and HerdSize data are placed.

4. Press **Submit**.

EpiTools calculates the required sample size (*n*) as 598 trees (Figure 5D). This can be interpreted as the minimum sample size that would enable us to be 95% confident of detecting citrus canker if it was present in ≥1% of trees in the population (about 230 infected trees). Other results in the table show the relationship between design prevalence and sample size—the lower the design prevalence, the higher the required sample size. This is because the lower the expected number of infested trees, the more difficult citrus canker will be to find, and so a larger survey will be necessary—a design prevalence of 0.005 (115 infected trees in the population) results in a sample size of 1197 trees out of the total population of >23,000 trees (5% surveyed), while a design prevalence of 0.001 (23 infected trees in the population) results in a sample size of 5990 trees, equivalent to surveying 26% of the population.

Note also that if test sensitivity were higher, the required sample size would be lower; for example, when test sensitivity is 1 (a perfect test), with desired herd sensitivity and design prevalence unchanged, $n = 299$ (the sample size halves).

To determine the 598 trees that are actually sampled from the total population of host trees, these steps should be followed:

1. Select **Survey Toolbox for livestock diseases and freedom in finite populations** from the EpiTools homepage (Figure 6A);

2. Select **Random sampling of animals from a list of owners** from the list of options that appears (Figure 6B);

3. Input the sample size determined earlier (**598**) (Figure 6C), select *fixed number*, and paste in the data for the sampling frame as indicated by Figure 6D. These data are contained in the columns *LocationID* and *PlantingSize* as discussed in Section 3.4.1.

4. Press **Submit**.

Results appear below the original input table, arranged in four columns (Figure 7). The data can be transferred into an Excel spreadsheet by scrolling down to the bottom of the page and clicking on **Detailed Results.** In either case, the first column of data under the heading *LocationID* contains the location identifier; the second column, *PlantingSize,* contains the number of trees at the particular location; the third column (*Number selected*) contains the number of trees that should be surveyed at a selected location; and the fourth column (*Individuals Selected*) lists which individual trees should be surveyed. So, for example, at location 15 there are 250 host trees, five of which should be surveyed. The random sampling

Results

Selected animals by owner

	LocationID	PlantingSize	Number selected	Individuals Selected
1	3	10	1	9
2	4	7000	185	69, 96, 171, 227, 230, 231, 248, 258, 274, 360, 403, 421, 431, 453, 481, 516, 536, 630, 674, 707, 728, 733, 753, 765 1534, 1651, 1690, 1817, 1949, 1962, 1997, 2067, 2079, 2086, 2127, 2203, 2214, 2286, 2292, 2300, 2311, 2315, 231 2974, 3088, 3146, 3243, 3276, 3290, 3312, 3337, 3372, 3390, 3410, 3446, 3472, 3492, 3499, 3525, 3537, 3538, 357 3884, 3933, 3960, 3992, 4019, 4041, 4051, 4110, 4215, 4228, 4283, 4289, 4342, 4357, 4396, 4410, 4419, 4449, 447 5052, 5065, 5075, 5109, 5137, 5185, 5201, 5241, 5315, 5319, 5351, 5400, 5440, 5474, 5493, 5511, 5551, 5588, 560 6108, 6114, 6138, 6278, 6376, 6405, 6441, 6471, 6518, 6521, 6625, 6708, 6
3	5	4000	93	5, 52, 74, 110, 173, 203, 316, 356, 377, 401, 403, 451, 467, 501, 504, 506, 538, 541, 578, 628, 638, 675, 678, 778, 81 1482, 1492, 1508, 1575, 1610, 1709, 1756, 1886, 1933, 1967, 2050, 2089, 2090, 2098, 2173, 2206, 2359, 2385, 240 3025, 3157, 3235, 3284, 3289, 3293, 3335, 3477, 3547, 3565, 3666, 3696, 3701, 372
4	6	2000	58	23, 43, 151, 197, 201, 233, 235, 249, 373, 409, 424, 486, 507, 510, 595, 648, 672, 695, 802, 883, 922, 928, 947, 958 1465, 1491, 1497, 1582, 1596, 1629, 1643, 1660, 1661, 1666, 1701, 1746, 1752, 1
5	7	1000	23	20, 131, 159, 260, 266, 273, 299, 303, 307, 402, 441, 490, 536, 540, 583,
6	8	980	24	12, 52, 162, 177, 275, 314, 352, 414, 417, 427, 540, 555, 561, 599, 616, 64
7	9	855	23	31, 155, 159, 161, 208, 214, 234, 242, 296, 330, 362, 369, 385, 392, 450,
8	10	660	18	14, 61, 124, 235, 243, 287, 303, 327, 367, 387, 397, 400, 48
9	11	500	17	4, 67, 94, 106, 146, 189, 217, 232, 263, 274, 296, 315, 3
10	12	400	13	32, 33, 56, 114, 179, 186, 191, 221, 223, 277, 2
11	13	400	17	75, 107, 167, 169, 188, 190, 218, 221, 237, 250, 276, 287,
12	14	400	9	4, 44, 73, 76, 104, 138, 263, 296, 30
13	15	250	5	121, 173, 191, 209, 225
14	16	200	2	111, 182

(A)

Figure 7. A view of a subset of the results from the one-stage survey using the citrus canker data

procedure identifies these as trees 121, 173, 191, 209, and 225 (Figure 7, Box A). When the on-ground surveying actually takes place, the survey officer should choose a logical starting point and select the trees in the order listed in the results. If it happens that a tree is missing at the location, a nearby tree should be selected. Alternatively, the survey officer could **randomly** choose five trees at the location by some other method.

Figure 8. The results from a one-stage survey for citrus canker showing the number of locations of a particular planting size (as number of trees per location) that will be sampled.

Results from the one-stage survey show that trees from a total of 72 locations should be surveyed. Recall that many of the locations containing host trees in the Northern Territory are in backyards, and it is important that these be captured by a surveillance strategy, along with trees in larger plantings (orchards). When results from the one-stage survey are broken down by number of trees per location, all planting sizes are represented (Figure 8): 18 locations (25%) have between one and five trees, 13 (18%) have between six and 10 trees; while seven (10%) have between 500 and 7000 trees.

It is also important that the surveillance strategy for citrus canker captures host trees in isolated locations. A breakdown of results by suburb/district reveals that of the 72 locations surveyed, 10 (14%) are from remote locations (Figure 9). Note that each time this random sampling procedure is repeated with the dataset, a new set of trees out of the 408 locations will result.
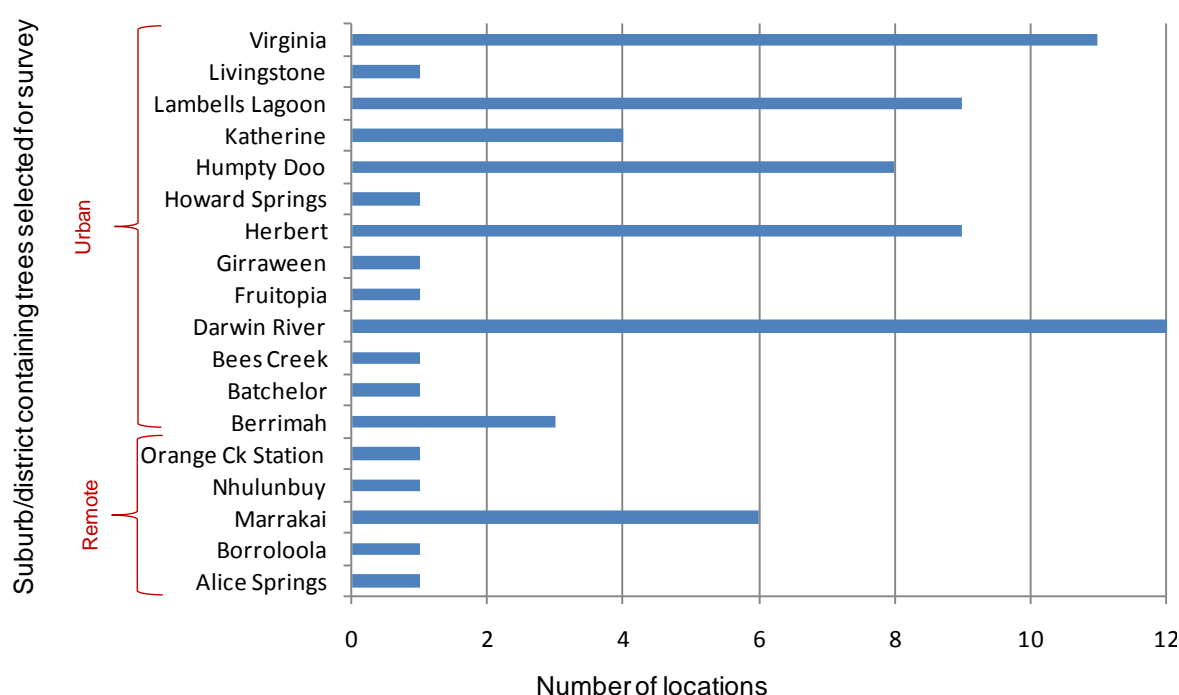
Figure 9. The results from a one-stage survey showing the various suburbs and districts where trees will be surveyed and number of locations in each district where surveying will occur.

A one-stage survey as described here can provide a very high level of confidence of detecting disease if it is present in the population at the specified design prevalence, at a reasonable cost. However, it will not provide an equivalent high level of confidence for the individual locations (orchards) sampled. If a high level of confidence for individual orchards is required, an alternative approach would be necessary; for example two-stage sampling (as described below), or a mix of twostage for orchards and one-stage for non-orchards. These methods

would provide much higher confidence for individual orchards selected but probably at greater cost.

3.4.3 A two-stage survey for citrus canker

A two-stage survey is one where sampling is undertaken at two levels: firstly, a sample of properties is selected from a list (called the 'sampling frame') of all properties with susceptible species; and second, a sample of trees is selected within each selected property. Two-stage sampling is particularly useful where a list of eligible properties is available but actual tree numbers for each property are unknown, so that a sampling frame of all trees cannot be constructed. As for one-stage sampling, simple random sampling (or equivalent) should be used at both stages for selecting properties and trees for testing. We mention alternative approaches to selecting the sample in Section 3.5.

There are several ways a two-stage survey could be undertaken for citrus canker in the Northern Territory:

1. a target location/orchard-level sensitivity (confidence of detecting infection if present at a specific location at the design prevalence) is specified. The target orchard sensitivity is often, but not necessarily, set at 95% to provide a very high level of confidence of detection for individual locations sampled. This desired orchard sensitivity is then used to calculate sample sizes for both stages: number of locations to sample, and number of trees to sample at each selected location; or

2. the cost of travelling to the location of the tree(s) and the cost of sampling/testing of each tree are used to determine the least-cost sample size. Depending on the ratio of costs between locations and individual trees, this can result in increased Stage 1 sample size (more locations), decreased Stage 2 sample size (fewer trees per farm) and reduced orchard sensitivity.[5]

Method 1

The following steps detail how to undertake a two-stage survey for citrus canker using Method 1:

Stage 1 – Determining the number of locations to sample

1. Select a target value for **location/orchard sensitivity** (often 95% but may be less) and overall **system sensitivity** desired from the survey (usually 95%).

---

[5] This method results in a trade-off between the average cost of sampling at a location (travel costs) and the number of samples taken at a location. If travel costs to each location are high, the result will be fewer locations in the sample, with more sampling at locations, and *vice versa*.

2. Select **Detection of a disease or demonstration of freedom** from the list of options contained in the EpiTools home page (Figure 10A)

3. Select **Sample size for demonstration of freedom in a finite population** from the list of options that subsequently appears (Figure 10B).

4. Insert values for **population size** (408 – the total number of locations), **test sensitivity** (this is now the location/orchard sensitivity specified in 1), **desired herd sensitivity** (this is the desired system sensitivity specified in 1) and herd- (orchard) level **design prevalence** (0.01 – the proportion of infected locations that you wish to be able to detect) into the appropriate place within the input box that should now appear (Figure 10C).

5. Press **Submit**
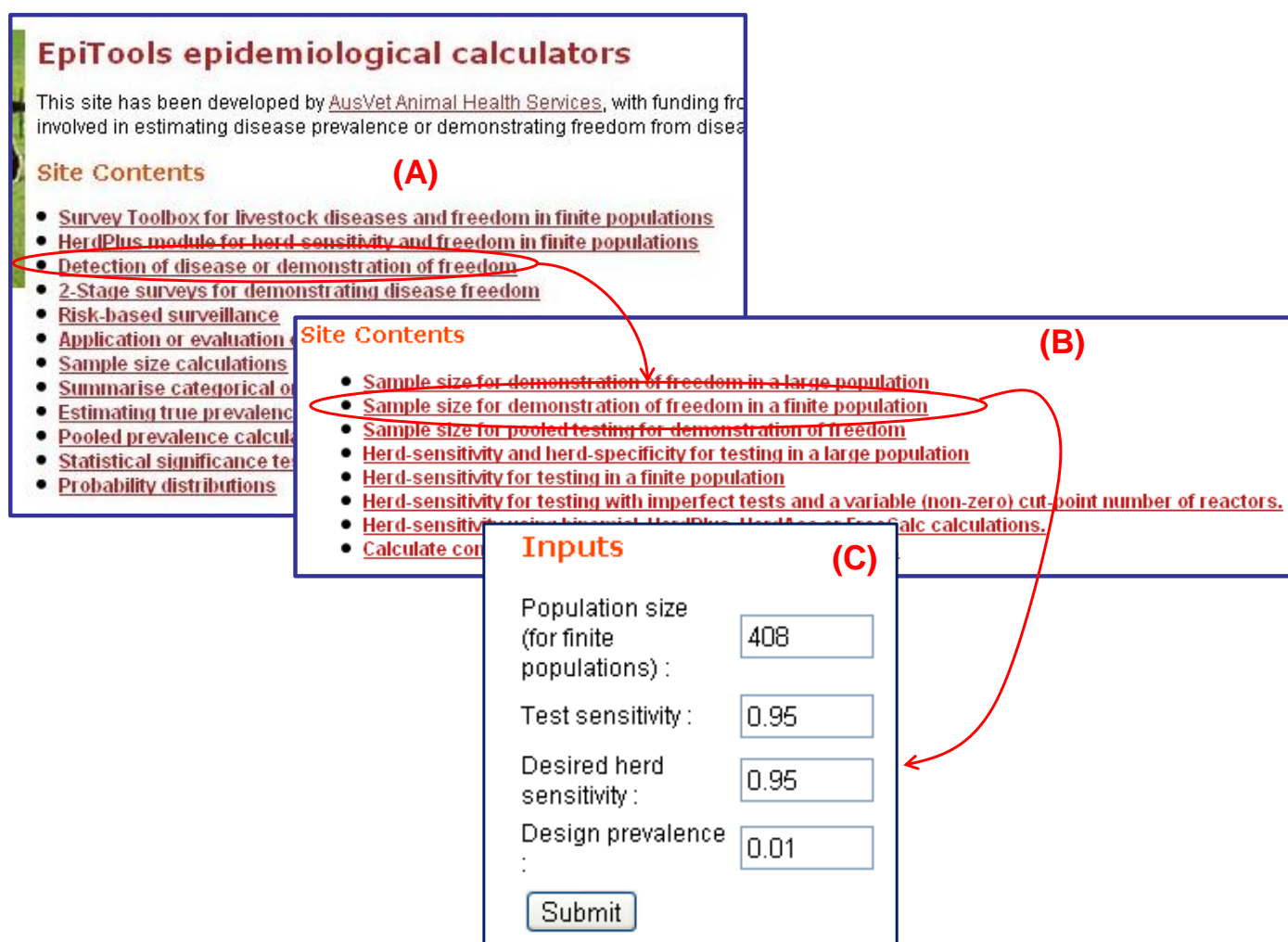
Figure 10. Screen views of steps involved in undertaking a two-stage survey using Method 1
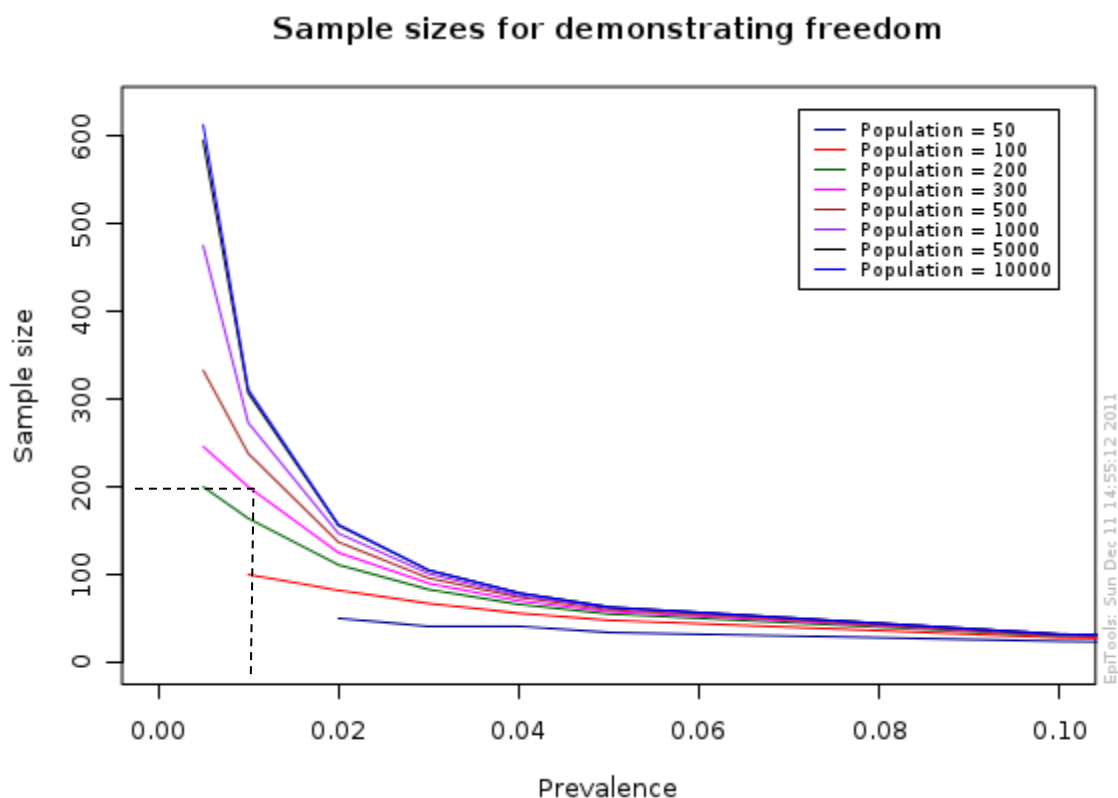
Figure 11. Sample sizes required to provide a probability of detecting disease of 95%, for various prevalence levels and population sizes for a test sensitivity of 0.95. The intersection of dashed lines represents the current scenario where the population is equivalent to the number of locations in the sampling frame (408), the (design) prevalence is 0.01 and resulting sample size is 194 (locations).

The required sample size for the specified level of confidence, design prevalence test sensitivity and population size, is given as 194. This is the number of locations that need to be surveyed. Since this is almost 50% of all locations, it means that many of those locations surveyed will be small properties. Additional results are provided as a table and graph indicating sample sizes required to provide the specified probability of detecting disease (95%), for various prevalence levels and population sizes for the specified test sensitivity (0.95) (Figure 11). Note that for a location/orchard sensitivity of 50%, it would not be possible to attain the desired level of confidence of detecting citrus canker if the design prevalence was reduced to 0.005, even if all locations were tested.

To find which 194 locations should be tested, either:

- Generate a list of 194 random numbers between 1 and 408 and select the corresponding locations from the list. Select **Survey Toolbox for livestock diseases and freedom in finite populations** in EpiTools, then **Generate a list of random numbers from a specified range or from a list**, enter the required sample size,

select sampling **without** replacement, select sample from **Specified range**, enter the minimum (1) and maximum (408) values, and click on **Submit**, or

- Generate a random selection of locations from the sampling frame developed previously (for one-stage sampling) using EpiTools. Select **Survey Toolbox for livestock diseases and freedom in finite populations** in EpiTools, then **Random sampling from a sampling frame**, enter the required sample size, select sampling **without** replacement, ignore stratification and sub-grouping, paste the sampling frame data into the input box, and click on **Submit**.
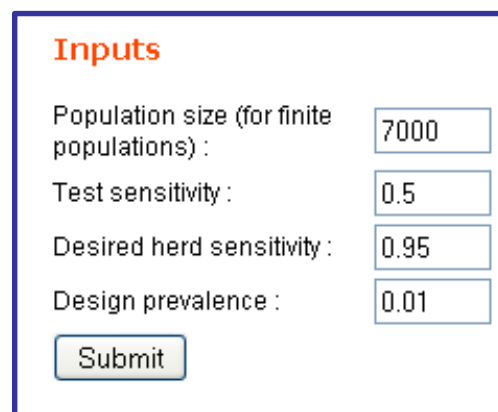
Once the list of selected locations has been generated, we can proceed to Stage 2.

Stage 2 – Determining the number of trees to sample at each location

For each selected location—which will have a different population size—repeat the process outlined below:

1. Select **Detection of a disease or demonstration of freedom** from the list of options contained in the EpiTools home page (Figure 10A)

2. Select **Sample size for demonstration of freedom in a finite population** from the list of options that subsequently appears (Figure 10B).

3. Insert values for **population size** (the number of trees at the selected location), **test sensitivity** (0.5, from the one-stage example and the previous national survey), **desired herd sensitivity** (this is the target location/orchard sensitivity specified in Stage 1) and **tree-level design prevalence** (0.01 – the proportion of trees infected at an individual location that you wish to be able to detect) into the appropriate place within the input box that should now appear (Figure 10C and Figure 12).

4. Press **Submit**

For an orchard with 7,000 trees, the input box in Step 3 would be filled out as shown in Figure 12. The required sample size is **587**. For 500 trees, the sample size is 451, while for 400 or fewer trees it is not possible to achieve 95% location sensitivity. For less than 100 trees, location sensitivity is equal to the test sensitivity of 50%. Considering this, it might be necessary to

Figure 12. An example of the input screen for Step 3 of Stage 2 of Method 1

increase the number of locations sampled to overcome the lack of location sensitivity.

In summary, the above approach works well where the primary sampling units (locations) are large (orchards), but not where there are numerous locations with small numbers of trees. In this situation, least-cost sampling (Method 2) is preferable, as this adjusts sample sizes at both levels to ensure that the target system sensitivity is achieved.

While this two-stage method might give more assurance of disease freedom on individual farms inspected, its high level of surveying means that it is usually a more expensive strategy.

Method 2

An alternative method for undertaking a two-stage survey is to consider the cost of undertaking the travel to the location and the actual cost of testing individual trees once at the location. In EpiTools, this is a two-stage survey for demonstration of freedom using least-cost sample size. This tool will require information to be provided on animal(tree)-level design prevalence, herd (orchard/backyard)-level design prevalence, test sensitivity, target system sensitivity, the relative testing cost per herd (orchard/backyard) and per animal (tree), and maximum sample size per herd.

The relevant values of prevalence, test sensitivity, and target system sensitivity for citrus canker are given in Table 1. The relative testing cost per orchard/backyard was calculated as $865, based on the average travel costs from Darwin to a range of locations, while the relative testing cost per tree was calculated from the average costs of labour required to take the sample and the average laboratory costs involved in culturing the sample. The option to provide a maximum sample size per herd is available in case a user wants to limit the numbers of hosts surveyed per property for logistic reasons—i.e. perhaps there is a maximum number of trees that can be inspected in a day, and inspectors will spend a maximum of one day per orchard. In the current scenario, a value of 1000 was chosen arbitrarily because in this example we do not want to limit the sample size per farm.

To undertake this type of two-stage survey, follow these steps:

1. Select **2-Stage surveys for demonstrating disease freedom** from the list of options contained on the EpiTools home page (Figure 13A);

2. Select **Calculate least-cost sample sizes and select herds for testing for 2-stage freedom survey where individual herd details are available** from the list of options that subsequently appears (Figure 13B).

3. Insert **animal-level design prevalence** (0.01) and select the proportion button,

4. Insert **herd-level design prevalence** (0.01) and select the **proportion** button,
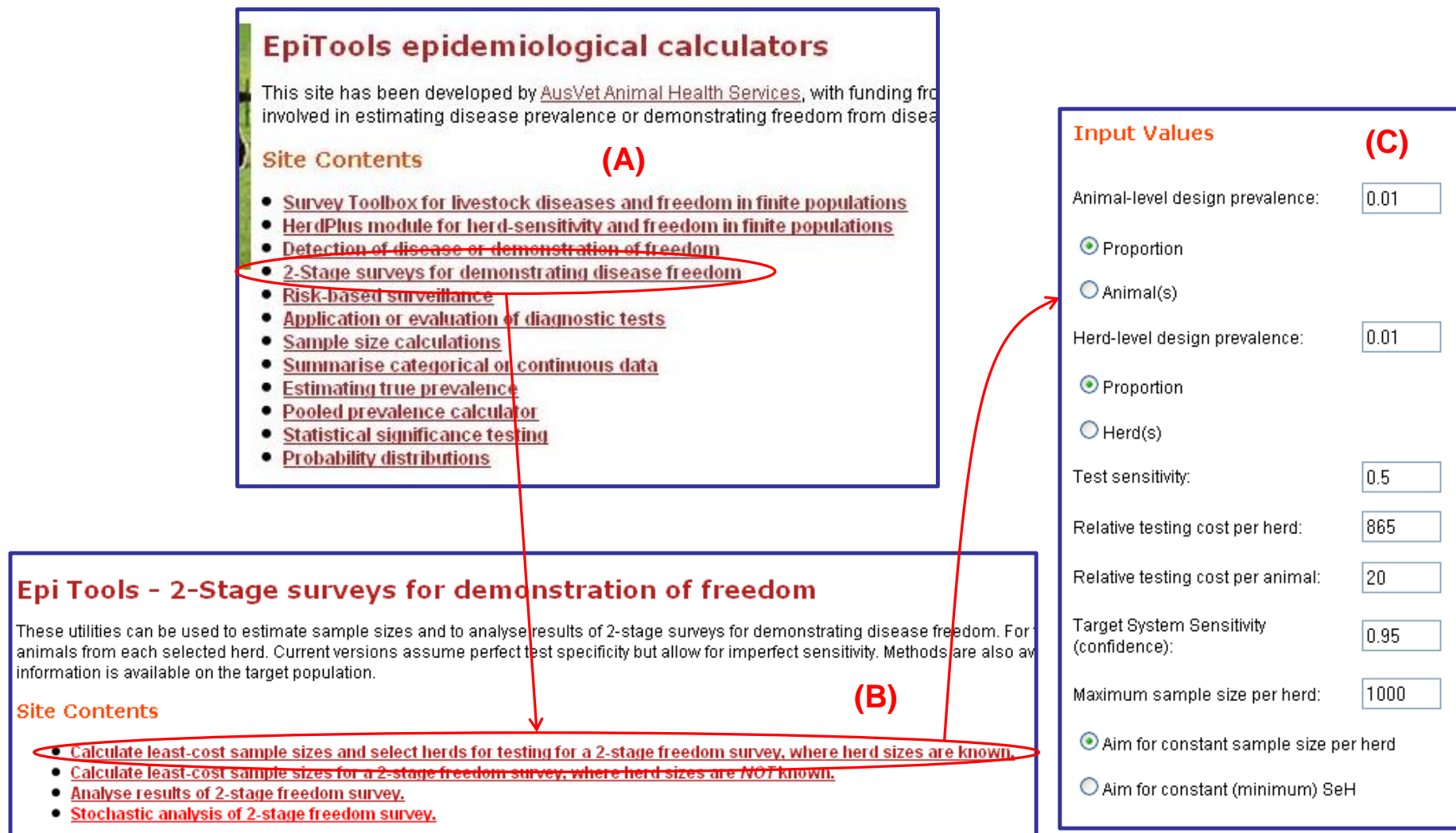
Figure 13. Screen views of steps involved in undertaking a two-stage survey using Method 2

1. Insert the **test sensitivity** (0.5), the **relative testing cost per herd** (orchard) ($865), and **per animal** (tree) ($20)

2. Insert **target System sensitivity** (0.95) and **maximum sample size per herd** (orchard/backyard) (1000) and select **Aim for constant sample size per herd** (so that those doing the inspections to do *x* trees at every location, or all trees if less than *x*) (Figure 13C).

3. Press **Submit**.

Results appear in summary form (Figure 14) and as five columns of data. The summary results indicate that 1885 trees will be tested across 327 locations with the survey strategy achieving an average herd sensitivity (probability of any selected location being identified as infected, if it is infected at the design prevalence) of 0.475 and a system sensitivity (probability of detecting infection if it is present in the population at the specified design prevalence values) of 0.95 (Figure 14). At each location, all trees are sampled up to a maximum of 28 (i.e. for locations with 28 or fewer trees all trees are sampled, for locations with more than 28 trees, only 28 randomly selected trees are sampled).

Results

Testing summary

Herds to test	327
Total sample size (animals)	1885
Samples per herd	28
Mean herd sensitivity (SeH)	0.4748
Achieved system sensitivity (SSe)	0.9509

Figure 14. Results from the two-stage survey with least-cost sample size

The output can be transferred into an Excel spreadsheet by scrolling down to the bottom of the page and clicking on the **Download** button. When downloaded into an Excel spreadsheet the data from the survey will start in Row 28. The first column of data contains numbers 1 to 327; the second column, *HerdID,* contains the location identifier; the third column, *HerdSize*, contains the number of trees at the particular location; the fourth column (*Sample size*) contains the number of trees that should be surveyed at a selected location; and the fifth column (*HerdSeH*) lists the herd (backyard/orchard) sensitivity for that location.
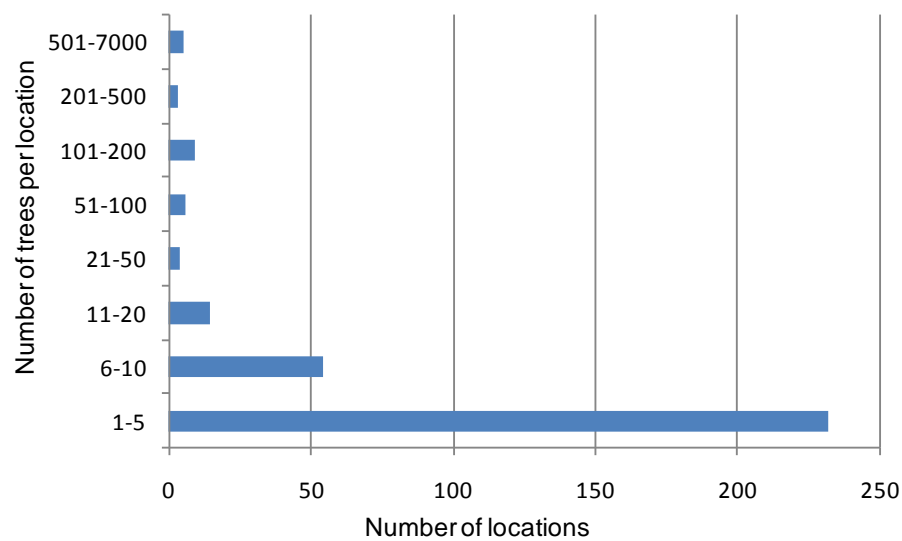
Figure 15. Results from a two-stage survey with least-cost sample size for citrus canker, showing the number of locations of a particular planting size that will be sampled.

Under the two-stage survey using least cost sample size, trees from a total of 327 different locations would be surveyed, with most locations having less than 10 host trees232 locations (71%) have between 1 and 5 trees, 54 (17%) have between 6 and 10 trees; while only five locations (2%) have between 501 and 7000 trees (Figure 15).
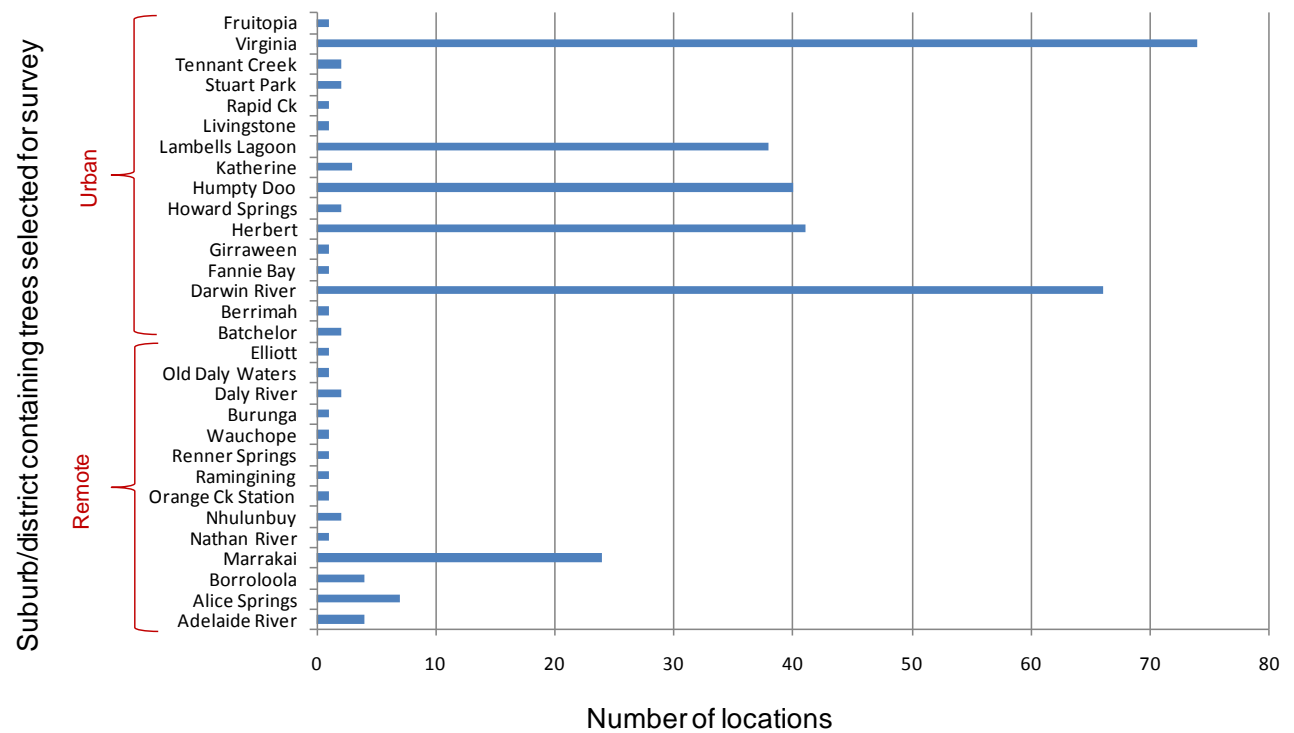
Figure 16. The results from a two-stage survey with least-cost sample size, showing the various suburbs and districts where trees will be surveyed and number of locations in each district actually surveyed

A breakdown of results by suburb/district reveals that of the 327 locations surveyed, 51 (16%) are from remote locations with the remainder surveyed in urban areas (Figure 16).

Note that if the cost of testing an individual tree is large, but the cost of collecting from other orchards/backyards is almost zero, the most efficient sample size will be one tree per farm.

While the sample size for each location is given in the output, a list of the individual trees that should be sampled at each location is not provided. For the current analysis, the output shows that this is not a problem for locations that have 28 trees or less—the output indicates that all trees at these locations should be sampled. For the 18 locations with more than 28 trees, the survey officer could randomly choose the trees to be surveyed, or for each location, EpiTools could be used to generate a list of random numbers (trees), in a similar way to that undertaken for a one-stage survey:

1. Select **Survey Toolbox for livestock diseases and freedom in finite populations** from the EpiTools home page (Figure 17A);

2. Select **Generate a list of random numbers from a specified range or from a list** from the options that appear (Figure 17B);

3. Input the **sample size** determined for a given location (e.g. 28) (Figure 17C),

4. Select the **Sampling without replacement** button, under the heading **Sampling with/without replacement?**

5. Select the **Specified range** button, under the heading **Random number source**,

6. Place a '1' in the box labeled **Enter minimum value for desired range** and place the number of trees at the location where the testing will take place (e.g. 100) in the box labeled **Enter maximum value for desired range** (Figure 17D). This maximum value may vary across locations.

7. Press **Submit**.

Results appear as a list of random numbers. In the case of the example used above, 28 different numbers will be given with values between one and 100. Once a list of trees has been generated, the survey officer would choose an appropriate starting point at the location and survey the individual trees listed. Note that each time this random sampling procedure is repeated, a new set of numbers will result.

Figure 17. Screen views of steps involved generating a list of random numbers to select trees that should be sampled at a given location

3.5. Discussion of different approaches

The main advantages of one-stage sampling are simplicity and per-unit efficiency compared with two-stage sampling (Cameron 1999); however, either approach is valid. The more labour-intensive of the approaches to two-stage sampling (Method 1) may cause problems unless applied carefully to ensure the target system sensitivity is achievable—i.e. that there are enough trees at each location to allow the appropriate level of testing as indicated by the required sample size.

If the aim is population—all citrus—freedom at minimum cost, the authors suggest either one-stage sampling or two-stage sampling with least-cost sample size (two-stage sampling, Method 2) as the best options. However, these approaches do not provide high confidence of freedom for individual orchards. If the requirement to achieve market access is for a high level of confidence of freedom in individual orchards, then sampling large orchards intensively becomes the best approach (two-stage sampling, Method 1) but at a significantly greater cost.

In deciding which option to pursue, we suggest the following: a decision should be made about appropriate design parameters for each approach; sample sizes should be calculated using these parameter values; and finally, comparisons between the various approaches can be made in terms of both the likely cost of surveying the given locations and the coverage of commercial operations. If the sensitivity for individual orchards is thought to be too low, the population could then be divided into commercial and non-commercial operations and a one-stage survey undertaken in non-commercial plantings, and two-stage survey (using Method 1) in commercial plantings with a specified high orchard sensitivity.

3.6. Field testing EpiTools

This case study shows initial field testing of EpiTools with data from a Northern Territory-wide survey for citrus canker. To facilitate further field testing of this tool, a workshop was held with NTDoR staff[6] where participants were guided through using the citrus canker data with EpiTools to design surveys for several proof-of-freedom scenarios. Participants were also able to get feedback on applying EpiTools to other plant (and animal) health problems and were encouraged to field-test the tool with other datasets. EpiTools has since been used with two additional plant-health survey problems:

i.   with myrtle rust (*Puccinia psidii s.l.*), to check the level of confidence of freedom from past surveys, and to develop rules of thumb about how many blocks should be surveyed within nurseries to maintain a particular level of confidence of freedom from myrtle rust at a low design prevalence, and

ii.   with cocoa-pod borer (*Conopomorpha cramerella*), to calculate a survey design that would allow the Northern Territory to have a particular level of confidence that if this pest were present at a low prevalence, it would have been found.

3.7. Next steps

This case study gives detailed instructions for how to design a proof-of-freedom survey for citrus canker, a highly contagious disease of citrus trees, using EpiTools, an existing set of web-based tools that are designed for this purpose. There is now a need for biosecurity managers to repeat the survey-design process described in this case study with other pests and diseases, and to report any problems to the authors during the process of developing an instruction manual.

---

[6] On 17 November 2011, Evan Sergeant delivered an EpiTools workshop with NTDoR staff at Berrimah Research Farm, Darwin. Participants in the workshop from the NTDoR were Stephen West, Brian Thistleton, David Hamilton, Gerry McMahon, Graham Schultz, Ian Miller, Jose Liberato, Mark Hearnden, Peter Saville, Susanne Fitzpatrick, Vicki Simlesa; from AusVet, Evan Sergeant, and from ACERA, Susie Hester.

In each of the designs, we have advocated the use of simple random sampling as a framework for sample selection. It is important to note that considerable improvements in efficiency may be realised by using other sources of information to formulate the design. For example, some of the orchards may be located in higher-risk areas, based on expert knowledge, or orchards may be in locations for which early detection provides greater benefits than other locations, or larger orchards may be more susceptible to invasion than smaller orchards, or even more likely to propagate an invasion that is underway. In these cases it is very useful to explore alternative sampling strategies within the one- and two-stage approaches; for example, stratification, variable probability sampling, and adaptive cluster sampling (Cochran 1977; Schreuder *et al.* 1993, Turk and Borkowski 2005). Using these more complicated options may well require different training and skills than are usually available to the biosecurity manager, and contracting of external expertise should be considered a useful option. Covering these designs is beyond the scope of EpiTools and, indeed, this report.

It is envisaged that using EpiTools to investigate citrus canker surveillance will be the first step in using and adapting this set of web-based tools for animal-health surveillance to solve a wide range of plant-based surveillance problems. For example, an additional issue that biosecurity managers face is that of determining how frequently surveys for a particular pest or disease should occur. There is some scope for EpiTools to assist in answering this question, but for many pests and diseases, survey frequency will be strongly related to climatic events (e.g. wet versus dry seasons, occurrence of cyclones, etc.) Survey frequency may also be related to the rate of spread of a pest or disease, the ability to detect the disease, and the cost of controlling the disease.[7] Given additional (time series) data for citrus canker, EpiTools could be used to demonstrate the calculation of confidence of freedom from this disease over multiple time periods. In this situation, EpiTools uses the concept of negative predictive value (NPV) to answer a question such as, what is the probability that citrus canker is not present, given that surveys have been undertaken for several years and nothing has been found?

3.8. Summary and Recommendations

Epitools can be used to design proof-of-freedom surveys in both the animal and plant-health fields, although the use of this tool has been poor in the latter. There are benefits to using EpiTools in the plant-health surveillance context via savings in time and thus cost—the user is directed to enter particular information that is necessary for the particular survey-design problem without needing to fully understand the underlying statistical formulas.

---

[7] These factors were found to be important for weeds when the ability to detect a species improves over time due to its increased spread, see Brown *et al.* (2004) and Harris *et al.* (2001).

The usefulness of EpiTools in designing structured surveys for plant-health problems has been clearly illustrated using the citrus canker application. This application could be replicated for other plant-health survey problems.[8] It is accessible to non-statisticians and is ready to be applied operationally. Use of EpiTools will improve the surveillance capacity in the jurisdictions and thus surveillance outcomes for Australia.

The use of EpiTools is recommended:

1. Where a structured survey is required to prove freedom in a plant-health context, to design surveys that will generate a required level of confidence (e.g. 95%) of detecting a disease/pest at or above a specified prevalence (e.g. 1%).

2. Where the budget for a structured survey is limited, to find the least-cost sample size that would be required in order to generate a particular level of confidence (e.g. 95%) of detecting a disease/pest at or above a specified prevalence (e.g. 1%).

In either case, survey designs could then be reviewed by a statistician if required.

---

[8] An instruction manual for using EpiTools to design proof-of-freedom surveys for plant-health problems is under development and will be available shortly.

4. Case Study 2: The eradication-monitoring tool

Susie Hester and Karen Herbert

4.1. Background

Eradicating pests and diseases can be a lengthy process. Therefore, eradication programmes need to be constantly reassessed to check progress towards the eradication objective and whether an alternative management objective may be preferable. Panetta and Lawes (2005) outline three criteria that can be used to evaluate weed eradication programmes: delimitation, containment, and extirpation. *Delimitation*—establishing the full spatial extent of a pest or disease incursion—is described by Panetta and Lawes (2005) as the fundamental criterion by which to evaluate an eradication programme. If an incursion is not delimited properly then its expansion may continue regardless of control that might be applied to known areas of incursion. *Containment* refers to preventing further spread of the incursion, and *extirpation* refers to the elimination of individual infestations within the delimitation area. Because containment can be difficult to prove, Panetta and Lawes (2005) suggest that checking conformity to the delimitation and extirpation criteria will be sufficient to assess progress toward the eradication objective.

Ideally, delimitation would be achieved as quickly as possible following detection of an incursion, because the invasive species continues to spread as searching takes place, increasing the probability of escape, the extent of the invasion, and the ultimate effort required to manage the invasion (Leung *et al.* 2010). In reality though, delimitation for most invasive pests and diseases does not occur rapidly, but is a gradual process.

To assist biosecurity managers to show the progress of extirpation/eradication attempts, a tool has been developed from a concept that combines delimitation and extirpation, as initially suggested by Panetta and Lawes (2007), and recently revised by Burgman *et al.* (submitted). The tool produces a graphical illustration of progress over time against both these measures.

4.2. The eradication-monitoring concept

To assess progress towards eradication, Panetta and Lawes (2007) developed the 'eradograph' concept, which is based on combining measures of the progress towards delimitation (D) and extirpation (E). The original eradograph concept has recently been revised by Burgman *et al.* (submitted) in order to correct incompatibilities in the units that were used to measure D and E. Whilst the revision retains the original intention of Panetta and Lawes (2007), it modifies the construction of the eradograph so that the axes of the

graph are interpretable and biologically meaningful. The following discussion draws heavily on the recent revision.

The variable *D* is derived from the ratio of new infested area and area searched, and the measure of extirpation, *Ex*, is derived from the 'monitoring profile' of infestations—the frequency distribution of time since the most recent detection of the weed at infested sites. The equations that are relevant to this case study are given in Appendix 2.

In the initial stages of an eradication programme, we might expect large values of *D* as surveillance activities are initiated and many new infestations are detected. Ideally, as the programme progresses, and given appropriate levels of searching, the value of *D* would reduce and eventually reach zero at some infested area, indicating successful delimitation. If treatment of the weeds that are detected is effective, and seed production is not allowed to occur, eventually no plants will be detected at infested sites although seeds or other propagules may still be present in the soil. When this is the case, these sites enter the 'monitoring' phase. The larger the number of sites in this phase, and the longer the amount of time spent in this phase, the smaller the value of *Ex*. Extirpation can be declared at individual sites when *Ex* = 0.

In contrast to the eradograph of Panetta and Lawes where the measure of delimitation was plotted against extirpation, in the revision, both measures are plotted against total area infested.

4.3. The eradication-monitoring tool

A test (beta) version of the revised eradograph tool has been developed in Microsoft Excel using the Visual Basic programming language.[9] Data from the branched broomrape eradication programme in South Australia, as given in Panetta and Lawes (2005, 2007), are used to illustrate use of the tool in sections 4.3.1 to 4.3.4 below. The tool is then field tested using data from the orange hawkweed eradication programme.

4.3.1 Initial Assumptions

As the initial step to using this tool, the user must enter the following information under the *Assumptions* heading on the worksheet titled *eradograph* (A in Figure 18):

- the year in which the eradication programme commenced (e.g. 1999);
- the latest year for which data is available (e.g. 2006); and

---

[9] To use the tool, users must ensure that macros are enabled in Excel. Data for branched broomrape are supplied in the test version of the tool.

- E_{max} - the average time (in years) since the final detection after which it may safely be concluded that the population has been extirpated (e.g. 7). This is often, but not always, equivalent to the seed longevity.
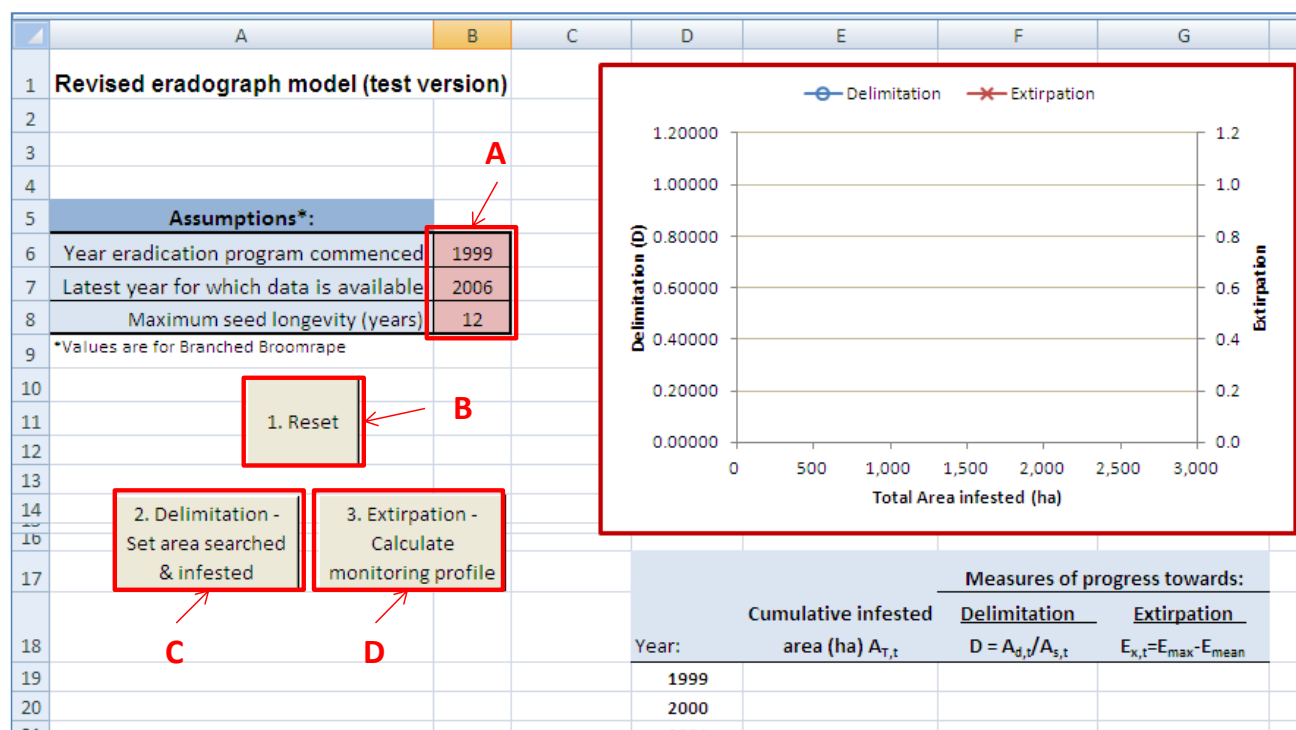
Figure 18. A screen view of the *eradograph* Sheet

4.3.2 Reset

Information from previous runs of the tool can be cleared by pressing the button called *1. Reset* (B in Figure 18).

4.3.3 Set Area Searched and Infested

In order to calculate progress towards delimitation (equation A7.1), the user must enter values for $A_{s,t}$ (area searched in a given year, t, of the programme) and $A_{d,t}$ (area of infestation newly detected in year t). This is done by pressing the button called *2. Delimitation - Set area searched & infested* (C in Figure 18), which takes the user to a new sheet (called *Area*) containing preset row and column headings, as well as columns containing the calendar year and which year in the eradication programme this particular year relates to. The cells in which values for $A_{s,t}$ and $A_{d,t}$ are to be placed are shown as A in Figure 19. Information provided in these cells is used by a formula in column F to automatically calculate $A_{T,t}$ (cumulative infested area in year t) as shown in B, Figure 19. Once the data have been entered, D (the delimitation measure) will be calculated by pressing the button **Calculate D**. Once this button is pressed, the user returns to the main

page. Values of D and $A_{T,t}$ for each year now appear in columns E and F, beginning at row 19.

Figure 19. Screen view of the *Area* sheet where data on area searched and newly detected area is entered.

4.3.4 Uploading the monitoring profile

The monitoring profile provides information on the length of time since a given infestation has been actively monitored, and is used in the calculation of the extirpation measure, E (see equation A7.2). To enter this information, on the main page press the button called ***3. Extirpation - Calculate monitoring profile*** (D in Figure 18)***.*** This opens a new sheet (called *mf*), enters appropriate row and column headings, and clears any existing information (Figure 20). Column headings are the calendar years of the eradication programme, and row headings are 'years since active' and will extend from zero to the number of years that the programme has been running. The user should enter values in each cell that represent, for the particular calendar year, the number of infestations that have been in the monitoring stage for the corresponding years since active. For example, the value of 205 in Row 4, column 5 (A in Figure 20) indicates that in the year 2000, there were a total of 205 sites that were being actively managed (so in the monitoring stage for 0 years). The value of 51 in Row 5, column D indicates that in the year 2000 there were a total of 51 sites where active treatment had not been applied for one year – i.e. these sites had been in the monitoring stage for one year.

Calculating $A_{e,t}$ from monitoring profiles

Site number by class and year

$A_{tot,t}$:

Number of sites:

Number of sites per year in each 'years since active' category:

Years since active:	1999	2000	2001	2002	2003	2004	2005	2006
0								
1								
2								
3								
4								
5								
6								
7								

Calculate Ex,t

Number of sites per year in each 'years since active' category:

Years since active:	1999	2000	2001	2002	2003	2004	2005	2006
0		205	194	52	273	66	122	145
1		51	155	180	24	243	46	95
2			39	140	107	14	191	32
3				28	107	105	12	177
4					24	105	93	6
5						19	93	76
6	A						23	88
7								13

Figure 20. Screen view of the *mf* sheet

Once information has been entered into the monitoring profile page, press the button called ***Calculate $E_{x,t}$***. Values for $E_{x,t}$ are subsequently calculated and entered next to the values for *D* on the *eradograph* worksheet. An eradograph is automatically plotted from these two columns of numbers (Figure 21). If the eradograph fails to plot check that the axis labelling is set to automatic (to do this, right click on each axis, select **Format axis**, then under **Axis options**, set each of the four options to **Auto**). Data labels showing years must be manually inserted into the graph each time a new eradograph is developed (to do this, right click on data points in graph so that they are all highlighted, then select **Add data labels**, and finally click on each label that appears and change it to the appropriate year). It may also be necessary to change the axis settings to improve the readability of the eradograph.

The original data from Panetta and Lawes were plotted in Figure 21 using the revised eradograph equations. The graph shows that both delimitation and extirpation are progressing, albeit slowly, towards the ideal of $D = 0$ and $E = 0$.
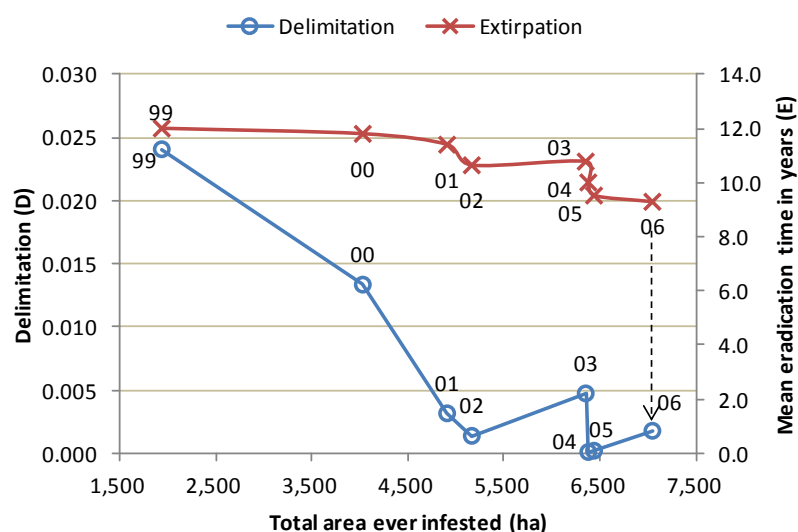
Figure 21. The branched broomrape data of Panetta and Pawes (2007) applied to (A) the revised eradograph, and (B) the original eradograph concept. Data labels are years.

In 2006 the value of *D* increased slightly, caused by a 900% increase in area of new infestations detected with only a 7% rise in area searched (data not shown). The data behind the extirpation line show that on average, 75% of infested sites move into the monitoring stage following treatment, but some of these sites revert back to active management even after several years in the monitoring stage (data not shown). Improved control methods would lead to a more rapid decrease in the extirpation line.

The arrows in Figure 21 demonstrate the ideal trajectory for each line. Trajectories towards the bottom right would indicate that management is effective. A delimitation curve that is heading towards the upper right quadrant suggests that increased search effort should be considered.

4.4. Application of the eradication-monitoring tool to orange hawkweed

The eradication-monitoring tool has been applied to the extirpation of orange hawkweed from the Australian Alps in Victoria.

4.4.1 Orange hawkweed

Orange hawkweed (*Hieracium aurantiacum)* is a perennial herb that grows up to 400 mm high with bright orange daisy flowers (Figure 22) (Johnson and Wright 2010). The weed is spread by both seeds and runners, and has the ability to establish in dense patches over large areas of the landscape (Morgan 2000). This weed has become widespread in New Zealand where it has replaced native vegetation and out-competed many varieties of pasture

Figure 22. Orange hawkweed (*Hieracium auranticum*)
Source: Johnson and Wright (2010)

species, resulting in decreased productivity of pastoral areas (CRC Weed Management 2003).

Orange hawkweed has become a weed in Tasmania's Central and Southern Highlands and around Hobart, after escaping from a garden planting sometime before 1963 (CRC Weed Management 2003). It was first reported as being naturalised in Victoria in 1999, in the Australian Alps (Thomas 2009). A naturalised infestation was first discovered in NSW in 2003, in the Kosciuszko National Park, and much of the alpine area of both NSW and Victoria has been identified as suitable for the further establishment of the weed (Johnson and Wright 2010). Since 2000, orange hawkweed has been the subject of an eradication programme in Victoria, and an eradication programme in NSW commenced in 2004.

Biosecurity managers responsible for allocating resources to orange hawkweed management would like to have a tool that can indicate whether their extensive survey and control efforts are leading towards eradication, and if they are not, how additional resources should be used to improve prospects for eradication.

One of the appealing features of the eradograph concept is that the spatial and temporal data required are usually routinely recorded during eradication programmes. The data required to undertake an eradograph analysis for orange hawkweed are:

- new detections of infested area over time (ha);
- area searched over time (ha);
- maximum seed longevity (yr);
- Status of infested sites over time (active or monitored, and years in monitoring).

4.4.2 The orange hawkweed data and its configuration for use in the eradograph tool

Data from the hawkweed eradication programme were supplied by Biosecurity Victoria and Parks Victoria. Parks Victoria maintains information about the programme in an Access database, with specific queries about data retrieved as Excel spreadsheets. There are currently 306 sites in the database. While the data required for use by the tool are routinely

recorded, some assumptions and further operations on the raw data were necessary as follows:

- Search and control activities take place between December and March, so records are in terms of financial years. To make the analysis easier, financial years were converted to calendar years, taking the first year recorded in the financial year as the calendar year; so, for example 2005/2006 becomes 2005.

- Data on area searched ($A_{s,t}$) were only available from 2004 onwards, but were considered unreliable until 2006.

- Data on new area infested ($A_{d,t}$) were available from 1999 although for some sites, actual area was not recorded. For these sites an area of 100 cm^2 was assumed on advice of personnel from Biosecurity Victoria.

- Information on the presence and absence of a species at a site was used to develop the monitoring profile in order to calculate E, the mean of the frequency distribution of the time since the most recent detection. In a small number of cases, the status of a site in a particular year was listed as *unknown.* Imputation based on reasonable presumption was used to give these sites a value of *present* or *absent*—for example, when the species was *absent* in the following year(s), *unknown* became *absent*, and when the species was present in the following year(s), *unknown* became *present.*

Through the process of applying the eradication-monitoring tool to orange hawkweed, it became apparent that deriving a monitoring profile can be a difficult task and additional instructions are needed if the tool is to be used by biosecurity managers. The following steps indicate how a monitoring profile was developed for orange hawkweed:

1. Columns containing *Site ID* and *site status* (P = present, A = Absent, U = unknown) for a given year were copied from the data into a new sheet. A sample of this data is presented in A, Figure 23.

2. Cells containing U were given a value of P or A depending on the status of surrounding cells.

3. Since the conversion of A and P into numeric values will occur to the right of the existing data, calendar years that correspond to the financial years were inserted to mark where this conversion will take place (B in Figure 23).

4. Numerical values will represent the 'years since active'. If a site is being actively managed in a particular year, its value will be 0. If it is being monitored in a particular year then its value should reflect the number of years since it was active. For

example, row 5, columns I to N contain P, P, A, P A, A—the site was actively managed for two years, then monitored for a year, actively managed for a year and monitored for two consecutive years—represented as 0, 0, 1, 0, 1, 2 in Row 5, columns W to AB (C in Figure 23). In the current case study, the large number of sites (306) meant that it was easiest to convert *A* and *P* into numerical values using 'IF statements'. An example of the IF statement for row 2, column W is given at D in Figure 24.

5. The final step is to calculate the number of sites containing a particular value for a particular year, and place the results in the monitoring profile (shown in Figure 20). The monitoring profile for orange hawkweed is not shown because this data remains confidential.
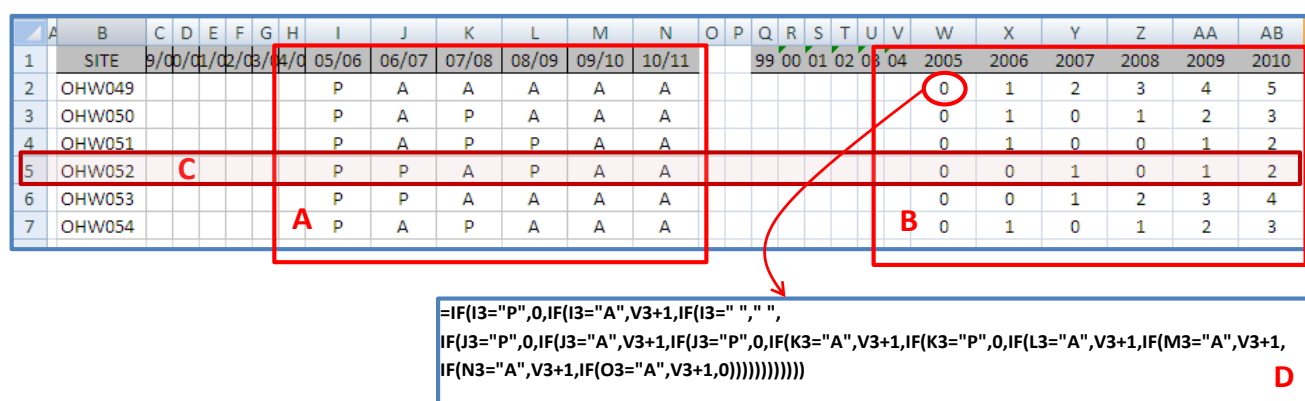
Figure 23. An example of the data showing status of a site: in terms of whether orange hawkweed was present (P) or absent (A), and how this information was converted to numerical values that showed the 'years since active' at a site.

4.4.3 A preliminary eradograph for orange hawkweed

Figure 24. Initial assumptions for orange hawkweed

In order to generate an eradograph for orange hawkweed, initial assumptions, data on area searched and infested, and the monitoring profile were all uploaded into the eradication-monitoring tool (see Section 4.3, steps 4.3.1 to 4.3.4) (assumptions shown in Figure 24).
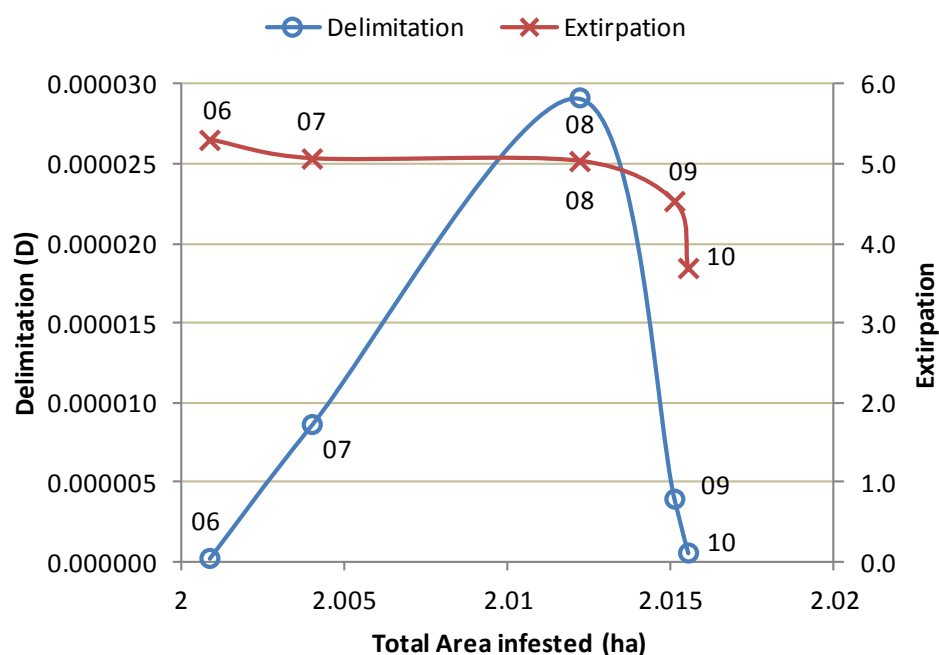
Figure 25. An eradograph showing progress in the extirpation of orange hawkweed, 2006 to 2010. Data labels are years.

The resulting eradograph is shown in Figure 25, for 2006 to 2010—the period for which the data are thought to be most reliable. Note that total area infested appears to have stabilised at 2.17 ha, while gross infested area (total area searched) was 791 hectares in 2010 (data not shown).

Overall it appears that progress towards eradication is being achieved, especially in the last few years—since 2008 (08/09) the line representing delimitation has fallen sharply, and the line representing extirpation has declined in every year since 2006.

The sharp rise in the delimitation line in 2008 indicates that progress was not being made towards delimitation in that year. Investigations are still continuing into why this sharp rise occurred.

The line representing extirpation gradually declines between 06 and 09, with an almost vertical fall between 09 and 10 indicating excellent progress towards extirpation more recently. The data behind the extirpation line show that on average 70% of sites move into the monitoring stage following treatment, with almost 100% of sites remaining in monitoring in the following years.

There is some uncertainty associated with E_{max} (time since last detection when eradication may be declared), and so the tool was used to calculate eradographs for E_{max} of 3 and 5

(years). The resulting eradographs are compared alongside the original in Figure 26. As E_{max} decreases the eradograph is positioned lower and lower on the y-axis. Progress towards eradication is more rapid the lower the value of E_{max}. For example, when $E_{max} = 3$, sites need only be monitored for three years before extirpation occurs, and the monitoring profile shows this has been easily achieved for many sites, compared to the situation when $E_{max} = 7$.
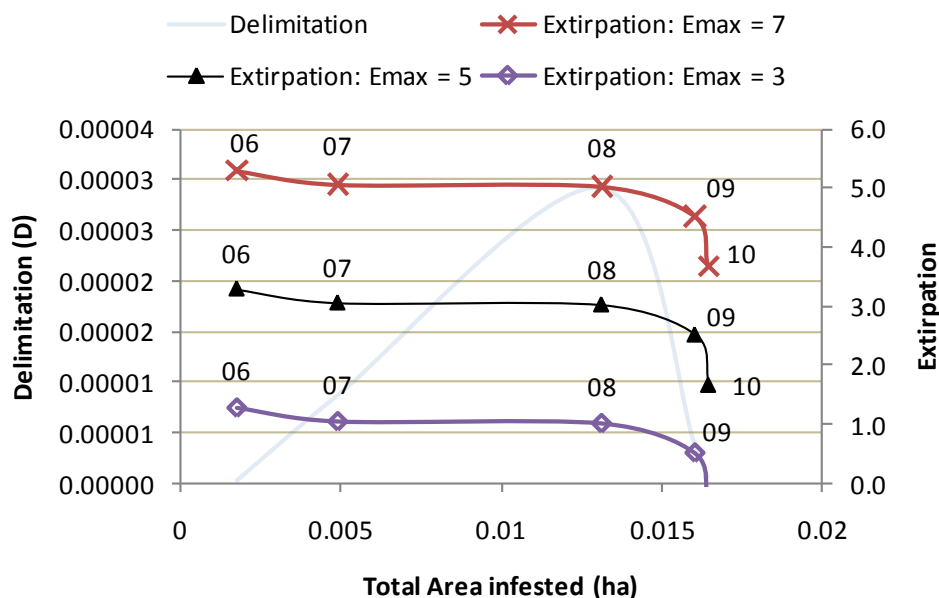
Figure 26. Eradographs for three different values of E_{max}.

The development of the monitoring profile raises the issue of when monitoring of sites should cease. An E_{max} of 7 implies that after a site has been monitored for seven years (no orange hawkweed plants detected during this time), there is no need to search the site in subsequent years. Interestingly, in 2010/11 there were approximately 70 sites that had been monitored for seven years (data not shown) and a couple of sites that were in their 11th year of monitoring[10].

For many sites, no initial area infested had been recorded when the weed was first discovered at a site, so an area of 100 cm^2 for each of those sites was assumed, in order to use those sites in determining the trajectory of delimitation. The effect of assuming a particular value for 'blank patch size' (BPS) on delimitation was investigated, using values for BPS of 1000 cm^2 and 10,000 cm^2. The outcome for delimitation from these two values is compared to the original delimitation curve (BPS = 100 cm^2) in Figure 27.

---

[10] Regan *et al*. (2006) explore the issue of when to declare eradication, and develop a rule of thumb where the optimal 'stopping time' is a trade-off between the cost of continued surveying and the cost of escape and damage if eradication is declared too soon.

Interestingly the path of decline in *D* between 2008 and 2010 remains unchanged for the three simulations because during these years there were few or no records where initial
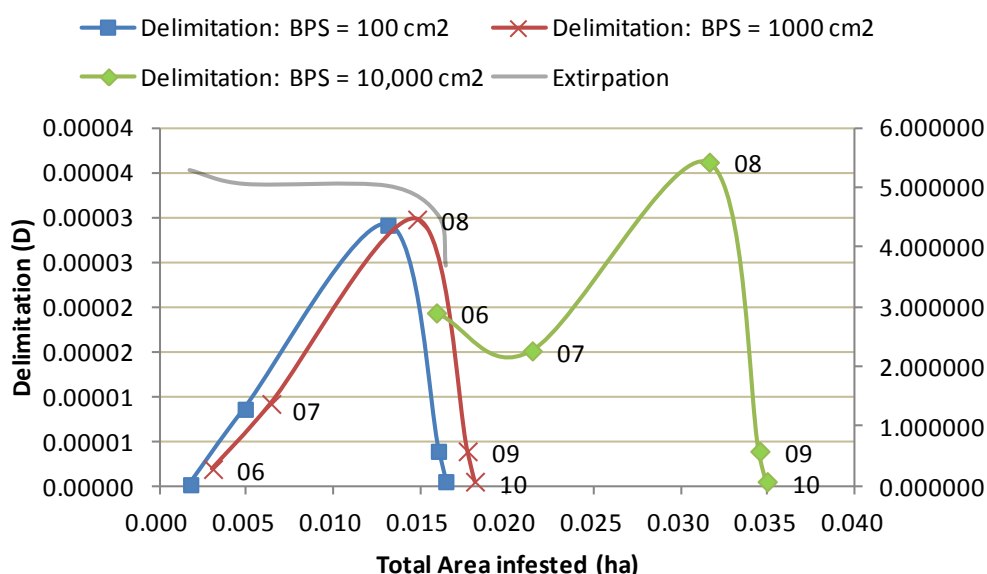
Figure 27 Eradographs for three different values of assumed blank patch size (BPS) where this value was not recorded at a site when an orange hawkweed was initially detected.

infested area was missing. The main effect of the different BPS values is on the total area infested in a given year—the larger the BPS value, the larger the total area infested, with the BPS of 10,000 cm^2 resulting in a much higher *D* value in 2008 and the emergence of a decrease in *D* between 2006 and 2007. It is thus important for managers to be confident in the value chosen for BPS—if the value falls within reasonable ranges of uncertainty then it is much less likely that there will be a substantial effect on the conclusion.

4.5. Field testing the eradication-monitoring tool

The eradograph tool was initially field tested with data from the orange hawkweed eradication programme in Victoria. The development of the monitoring profile was identified as a confusing task, so detailed instructions are now provided on how to proceed with the task of constructing monitoring profiles.

The eradication-monitoring tool was recently presented to a meeting of Victorian State Prohibited Weed (SPW) officers,[11] and a request made for the tool to be field tested on SPWs and for any problems with the tool to be reported to the authors. As a result the tool was applied to King devil hawkweed (*Hieracium piloselloides*) and mesquite (*Prosopis* spp.) incursions in Victoria (eradographs not shown due to confidentiality). So far, only one minor

---

[11] On 27 October 2011, Susie Hester met with Victorian SPW Officers and Victoria DPI colleagues at Attwood, Melbourne, to discuss the eradication-monitoring tool. Present at the meeting were Neil Smith, Karen Herbert, Sarah Partington, Erin Cox, Michael Hansford, Sarah Brunel, and Emily Hart

problem with the tool has been reported—the lines for delimitation and extirpation did not show on the graph as expected. The problem was fixed by setting the axis labeling to automatic. Both the instruction manual and the tool were modified to ensure this will not be a problem in future applications of the tool.

4.6. Next steps

This tool could be usefully extended to scenario analysis of future changes in management strategy. For example, when an eradication programme appears not to be progressing towards its eradication goal, it is envisaged that the tool be used to explore the cost of increasing the control effort and/or the search effort required to improve progress. In order to undertake this extension, information would be needed on the relationship between search effort and detection of plants, and between time spent controlling and effectiveness of control methods.[12]

4.7. Summary and recommendations

This case study gives detailed instructions on how to use data from a weed-eradication programme in a spreadsheet-based eradication-monitoring tool in order to show the progress of extirpation/eradication attempts, depicted as an eradograph. The tool was developed from a concept that combines delimitation and extirpation, as initially suggested by Panetta and Lawes (2005) and recently revised by Burgman *et al.* (submitted).

The eradication-monitoring tool is accessible to biosecurity managers and is ready to be applied operationally. Use of this tool provides a convenient method for evaluating progress in an eradication programme as a basis for making sound decisions on the future delivery of such programmes:

- to justify continued investment if the eradograph demonstrates that progress towards eradication is good, or

- consideration of increased expenditure, redesign of the surveillance and control activities, or changing the objective altogether, if the eradograph demonstrates that progress towards eradication is poor.

The use of the eradication monitoring tool, incorporating the eradograph, is recommended:

1. Where an objective ongoing measure of the progress of a weed eradication programme is needed to assist decision making on future delivery of the programme

---

[12] Hester *et al.* (2011) provides an initial attempt at modeling these relationships.

5. References

ABARE (Australian Bureau of Agricultural and Resource Economics) (2009). *Australian Commodity Statistics 2009*, Canberra.

Animal Health Australia (2008a). Summary Document (Version 3.1) Australian Veterinary Emergency Plan (AUSVETPLAN), Edition 3, Primary Industries Ministerial Council, Canberra, ACT.

AHA (Animal Health Australia) (2008b). *National TSE Surveillance Program (NTSESP) National Guidelines for Field Operations*: July 2009 to June 2010.

Brown, JA, Harris, S, and Timmins, S (2004). 'Estimating the maximum interval between repeat surveys'. *Austral Ecology* 29 631−636.

Burgman, MA McCarthy, MA, Robinson, A, and Hester, SM (Submitted). 'An extension of the Panetta-Lawes eradication graph'. Submitted to *Weed Science*.

Cameron, AR (1999). Survey Toolbox for Livestock Diseases – A practical manual and software package for active surveillance in developing countries, ACIAR Monograph 54, vii + 330p.

Cochran, WG (1977). *Sampling Techniques*, 3rd Edition. Wiley.

CRC Weed Management (2003). Weed Management Guide: Orange hawkweed – *Hieracium aurantiacum*. http://www.weeds.gov.au/publications/guidelines/alert/pubs/h-aurantiacum.pdf

DAFF (Department of Agriculture, Fisheries and Forestry) (2009) Plant Protection News-February 2009. http://www.daff.gov.au/animal-plant-health/plant/publications/ppn/february2009#erad

FAO (2002). *Guidelines for surveillance of specific pests: Xanthomonas axonopodis pv. citri (Citrus canker).* (Food and Agriculture Organisation of the United Nations: Rome)

Harris, S, Brown, J, and Timmons, S (2001). 'Weed surveillance – how often to search?' *Science for Conservation* 175.

Hester, S. Cacho, O. and Panetta, D. (2011). 'A tool for monitoring progress towards eradication', paper contributed to the 55th Annual Conference of the Australian Agricultural and Resource Economics Society, Melbourne, 8-11 February, 2011.

Hester, S, Hauser, C, Kean, J, Walshe, T, and Robinson, A (2010). 'Post-border surveillance techniques: review, synthesis and deployment', Draft Milestone Report, ACERA 1004, University of Melbourne.

Johnson, B, and Wright, G (2010). *Orange Hawkweed Control Program, Annual Report 2009/10.* Unpublished Report. DECCW, New South National Wales Parks and Wildlife Service.

Leung, B, Cacho, O and Spring, D (2010). 'Searching for non-indigenous species: rapidly delimiting the invasion boundary'. *Diversity and Distributions*. 16(3): 451-460.

Morgan JW (2000) Orange hawkweed *Hieracium aurantiacum* L.; a new naturalised species in alpine Australia. *The Victorian Naturalist*, 117(2).

Moore, JL (2011). 'Contain or eradicate? Optimizing the management goal for Australian acacia invasions in the face of uncertainty'. *Diversity and Distributions*, DOI: 10.1111/j.1472-4642.2011.00809.x

OCPPO (Office of the Chief Plant Protection Officer) (2004). *Surveillance for Citrus Canker to Achieve Pest Free Area Status in Regions Outside of Queensland*. Unpublished Draft Report prepared by the OCPPO in consultation with a technical surveillance working group for consideration by Domestic Quarantine Marker Access Working Group

Panetta, FD, and Lawes, R (2005). 'Evaluation of weed eradication programs: the delimitation of extent'. *Diversity and Distributions* 11, 435–442.

Panetta, FD, and Lawes, R (2007). 'Evaluation of the Australian branched broomrape (*Orobanche ramosa*) eradication program'. *Weed Science* 55, 644-651.

PHA (Plant Health Australia) (2009). Industry Biosecurity Plan for the citrus industry (Version 2.0). Plant Health Australia, Canberra, ACT. http://www.planthealthaustralia.com.au/

PHA (Plant Health Australia) (2006). *National mango industry biosecurity plan*. Deakin, ACT http://www.planthealthaustralia.com.au/.

PHA (2007). *National cherry industry biosecurity plan*. Deakin, ACT <http://www.planthealthaustralia.com.au/>

Regan, T, McCarthy, MA, Baxter, PWJ, Panetta, FD, Possingham, HP (2006). 'Optimal eradication: when to stop looking for an invasive plant, *Ecology Letters* 9, 759-766.

Rout, TM, Spring, D, Runge, MC, Robinson, A, and Walshe, T, (2010) 'Time preference and value of information in the context of estimating consequence impacts'. ACERA 1002 Research Report, University of Melbourne, 92pp.

Thomas. E (unpublished 2009). *Victorian Alps Annual Hawkweeds Control Program Draft Report*. Mt Beauty, Parks Victoria.

Schreuder, HT, Gregoire, TG, and Wood, GB (1993). *Sampling methods for multiresource forest inventory*. Wiley, New York.

Sergeant, ESG (2009). *Epitools epidemiological calculators*. AusVet Animal Health Services and Australian Biosecurity Cooperative Research Centre for Emerging Infectious Disease. http://epitools.ausvet.com.au.

Turk, P, and Borkowski, JJ (2005). A review of adaptive cluster sampling: 1990-2003. *Environmental and Ecological Statistics* 12, 55–94.

6. Glossary

Area infested: the total area actually occupied by an invasive pest or disease.

Area searched: the area over which search effort is applied with the aim of detecting an invasive pest or disease.

Containment: preventing the further spread of the incursion.

Delimitation: the process of determining the spatial extent of a pest or disease incursion.

Design prevalence: the pre-survey estimate of likely true (actual) prevalence. It is the plant- or animal-level prevalence of a pest or disease to be used in calculating sample size. It is expected that the design prevalence (and true prevalence) are near zero when claiming area freedom is the objective.

Eradograph: a concept, which is based on combining measures of the progress towards delimitation (*D*) and extirpation (*E*) as a graphical illustration, originally developed by Panetta and Lawes (2007) and recently extended by Burgman *et al.* (submitted).

Extirpation: the elimination of individual infestations within the delimitation area.

Monitoring profile: the frequency distribution of time since the most recent detection of the weed at infested sites.

Population: the entity for which one wishes to estimate quantities of interest or answer questions about.

Sampling: the process of selecting a sample from the population. Appropriate sampling methods for choosing the sample in the proof of freedom context are simple random sampling, stratified random sampling, systematic sampling, and sampling using random spatial coordinates.

Sensitivity: the proportion of truly positive units that are correctly identified as positive by a test.

Specificity: the proportion of truly negative units that are correctly identified as negative by a test.

Surveillance: the collection, collation, analysis, interpretation, and timely dissemination of information on the presence, distribution, or prevalence of pests or diseases, and the plants or animals that they affect.

Survey: an investigation, in which information is systematically collected, usually carried out on a sample of a defined group or area, within a defined time period.

Tool: a readily deployable rule of thumb, formula, simulation model or piece of software.

True (actual) prevalence: the number of sick animals or infested trees at a single point in time, as a proportion of the total population at risk at that time

7. Appendix 1: Important formulae for surveillance

Freedom and risk-based surveillance

Herd/system sensitivity and Sample size

Binomial	$SeH = 1 - (1 - Se \times P^*)^n$	Assumes sampling with replacement or sample small (<10%) relative to population
	$SSe = 1 - \prod (1 - SeH_i \times P^*)$ where SeH varies among herds	
	$n = \log(1 - SeH)/\log(1 - Se \times P^*)$	
Hypergeometric approximation	$SeH = 1 - (1 - Se \times n/N)^{P^* \times N}$	Sampling without replacement and where sample size is large relative to population
	$SeH = 1 - (1 - Se \times n/N)^d$	
	$SSe = 1 - (1 - ASeH \times h/H)^D$	
	$n = (N/Se)*(1 - (1 - SeH)^{1/(P^* \times N)})$	
Exact	$SeH = 1 - (1 - Se)^d$	Where the entire population is sampled
	$SSe = 1 - (1 - ASeH)^D$	

Negative Predictive Value (or confidence of freedom)

$NPV = (1 - PrInf)/(1 - PrInf \times SSe)$

Assuming perfect (100%) test/system specificity.

Revising confidence of freedom in successive time periods

$PrInf_t = PostPInf_{t-1} + PIntro_t - PostPInf_{t-1} \times PIntro_t$

Where $PostPInf_{t-1} = (1 - NPV_{t-1})$

Equilibrium $NPV = (1 - (PIntro / SSe)) / (1 - PIntro)$
(maximum or minimum value possible for NPV for given combinations of SSe and Pintro)

Combining test sensitivities in series
(For example in a diagnostic process with multiple steps)

$Se_{combined} = \prod Se_i$

Combining component sensitivities in parallel, assuming independence

Calculates SSe from multiple components, assuming independence (no overlap between units sampled) between components, for example different compartments or different herds/flocks represented in the surveillance system.

$$SSe = 1 - \prod (1 - CSe_i)$$

Updating herd sensitivities between components where there is overlap

This assumes no independence between components, for example where the same herds or flocks are represented in multiple surveillance system components. The probability of infection for each herd is adjusted between components and resulting component sensitivities are then combined as for assuming independence. For this example binomial calculations are used, but hypergeometric or exact could also be use dif appropriate:

Method 1: Adjusting effective probability of infection between components:

1. Calculate SeH for each herd $[SeH_1 = 1 - (1 - Se_1 \times P^*)^n]$ for each component.

2. Calculate posterior confidence of freedom and hence posterior probability of infection for each herd for the first component (component order is a matter of convenience):

    $NPV_h = (1 - P^*)/(1 - P^*_h \times SeH)$ where P^*_h is the herd-level design prevalence

    $PostPInf_h = (1 - NPV_h)$

3. Calculate probability that each herd has a negative test result and hence component sensitivity for first component:

    $P(Neg) = 1 - P^*_h \times SeH$

    $CSe = 1 - \prod (P(Neg))$

4. Calculate P(Neg) for each herd and CSe for the second component after substituting $PostPInf_h$ instead of P* or EPI in formula:

    $P(Neg) = 1 - PostPInf_h \times SeH$

    $CSe = 1 - \prod (P(Neg))$

5. Repeat for as many components as necessary

6. Herds start with P* (or EPI) at the first component in which they appear and then get updated as necessary

7. When all component sensitivities have been calculated, calculate overall system sensitivity (probability that one or more components will yield a positive result if the population is infected at the design prevalence), using independence formula.

    $$SSe = 1 - \prod (1 - CSe_i)$$

Method 2: Aggregating data between components:

An alternative (often simpler) approach is to aggregate the data for each herd to calculate single SeH values and then combine these values to calculate overall system sensitivity:

$$SeH = 1 - \prod(1 - P^*_a \times Se_i)^{\wedge}n_i)$$

For where Se_i and n_i are test sensitivity and sample size for each of the i components in the surveillance system.

Risk-based surveillance

Adjusted risk and effective probability of infection

$$AR_{low} = 1/(RR \times PPrH + PPrL)$$
$$AR_{high} = RR \times AR_{low}$$

$$EPI = P^* \times AR \text{ (for respective risk categories)}$$
EPI > 1 is invalid.

System sensitivity for simple, 1-stage, no risk factors, one factor affecting sensitivity

$$SSe = 1 - (1 - P^* \times Se_{high})^{n(h)} \times (1 - P^* \times Se_{low})^{n(l)}$$

n(h) and n(l) are sample sizes for high and low sensitivity groups, respectively.

Sample size for simple, 1-stage, one risk factor, constant sensitivity

$$USe = EPI_{high} \times Se \times SPr + EPI_{low} \times Se \times (1 - SPr)$$

$$n = \log(1 - SSe)/\log(1 - USe)$$

Se is average sensitivity in units tested (animals or herds, depending on level of analysis).

USe is unit sensitivity = probability of a single randomly selected animal from the sample being positive, given the population is infected at the design prevalence.

SPr is the proposed sample proportion from the high-risk sub-population.

System sensitivity for simple, 1-stage, one risk factor, one factor affecting sensitivity

$$SSe = 1 - (1 - EPI_{high} \times Se_{high})^{n(hh)} \times (1 - EPI_{high} \times Se_{low})^{n(hl)} \times$$
$$(1 - EPI_{low} \times Se_{high})^{n(lh)} \times (1 - EPI_{low} \times Se_{low})^{n(ll)}$$

n(hh), n(hl), n(lh) and n(ll) are sample sizes for high risk & high sensitivity, high risk & low sensitivity, low risk & high sensitivity and low risk & low sensitivity groups, respectively.

Sample size for simple, 1-stage, one risk factor, one factor affecting sensitivity

$LRSe = SPr_{LH} \times Se_{high} + (1 - SPr_{LH}) \times Se_{low}$

$HRSe = SPr_{HH} \times Se_{high} + (1 - SPr_{HH}) \times Se_{low}$

$USe = EPI_{high} \times HRSe \times SPr + EPI_{low} \times LRSe \times (1 - SPr)$

$n = \log(1 - SSe)/\log(1 - USe)$

LRSe, HRSe are weighted average sensitivity in low and high risk samples respectively.

USe is unit sensitivity = probability of a single randomly selected animal from the sample being positive, given the population is infected at the design prevalence.

SPr, SPrLH , SPrHH are proposed sample proportions from the high-risk sub-population, high sensitivity group in high-risk sub-population and high sensitivity group in low-risk sub-population respectively.

4

Key:

Abbreviation/symbol	Meaning
n, N	Sample size and corresponding population size (animal level)
h, H	Sample size and corresponding population size (herd level)
d, D	Number of diseased elements in population (animal and herd levels respectively)
t	Time period
P^*	Design prevalence at appropriate (animal or herd) level, can substitute EPI for the relevant risk group in risk-based surveillance
Se	Test sensitivity (herd-level sensitivity when calculating system sensitivity)
SeH	Herd/flock sensitivity (= SSe when calculating system sensitivity)
SeH_i	Herd/flock sensitivity for the i-th herd/flock
ASeH	Average herd sensitivity across herds sampled
SSe	System sensitivity (or herd sensitivity when working at herd level)
CSe_i	Component sensitivity for the i-th surveillance system component
PrInf	Prior probability of being infected = 1 – prior confidence of freedom
PostPInf	Posterior probability of being infected = 1 – posterior confidence of freedom (NPV)
NPV	Negative Predictive Value = posterior confidence of freedom
RR	Relative risk
AR	Adjusted risk
PPrH, PPrL	Population proportions in high and low risk groups, respectively
EPI, EPI_{high}, EPI_{low}	Effective probability of infection and EPI in high and low risk groups. Probabilities of infection after adjusting design prevalence for group relative risks
Se_{high}, Se_{low}	Sensitivity in high and low risk groups, respectively. May be test (animal) sensitivity or herd-sensitivity, depending on level at which being calculated.
USe	Unit sensitivity = probability of a single randomly selected animal from the surveillance sample being positive, given the population is infected at the design prevalence.
log	Natural logarithm
P(Neg)	Probability of a negative test result (= 1 – sensitivity × probability of infection)
$\prod$	Product of a series of elements represented in parentheses following

Prevalence estimation

Apparent or seroprevalence
(assumes perfect test sensitivity and specificity)

Estimated prevalence: $P = x/n$

Asymptotic (normal approximation) confidence intervals:

$$CI = P \pm Z\sqrt{((P(1 - P)/n)}$$

Alternative (binomial, Wilson binomial) CI methods usually better, particularly as P approaches 0 or 100%.

Sample size: $n = (Z^2 \times P(1 - P))/e^2$

Assumes a large population. Where expected sample size is large (10%) relative to populations size use following adjustment:

$$n_{adj} = (N \times n)/(N + n)$$

Estimated true prevalence
(allows adjustment for imperfect sensitivity and specificity)

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

Note: Method fails when Se + Sp = 1 due to division by 0.
TP may be negative if AP + Sp < 1 (Sp estimate is lower than suggested by the results).

Asymptotic (normal approximation) confidence intervals assuming known sensitivity and specificity :

$$CI = TP \pm Z\sqrt{[AP(1 - AP)/(n \times (Se + Sp - 1)^2)]}$$

Assumes Se and Sp known exactly (no uncertainty).
Lower CI may be <0 if TP is close to 0.

Sample size:

$$n = (Z/e)^2 \times (Se \times TP + (1 - Sp) \times (1 - TP)) \times (1 - Se \times TP - (1 - Sp) \times (1 - TP))/(Se + Sp - 1)^2$$

Asymptotic (normal approximation) confidence intervals assuming uncertain sensitivity and specificity :

$$CI = TP \pm Z\sqrt{[AP \times (1-AP)/(n \times (Se + Sp - 1)^2) +}$$
$$(Se \times (1-Se) \times TP^2)/(M \times (Se + Sp - 1)^2) +$$
$$(Sp \times (1-Sp)*(1-TP)^2)/(R*(Se + Sp - 1)^2)]$$

Key:

Abbreviation/symbol	Meaning
n, N	Sample size and corresponding population size (animal level)
P	Observed or expected prevalence (proportion)
x	Number of units with the characteristic of interest
Z	Z distribution value corresponding to desired confidence level Z = 1.96 for 95%, 2.58 for 99% and 1.64 for 90%
e	Desired precision of estimate (± relative to estimate). Confidence interval width = 2e
n_{adj}	Sample size adjusted for small population
TP	True prevalence estimate
AP	Apparent prevalence estimate
Se, Sp	Sensitivity and specificity of the test used
CI	Confidence interval
M	Sample size for estimating test sensitivity
R	Sample size for estimating test specificity

8. Appendix 2: Technical notes on the revised eradograph

The following equations and discussion of the eradograph are drawn from Burgman *et al*. (submitted) which modifies the original concept and equations of Panetta and Lawes (2007).

Delimitation

The delimitation measure, *D*, is derived from the ratio of new infested area to area searched:

$$D_t = {A_{d,t}} \Big/ {A_{s,t}} \qquad\qquad A_{s,\,t} \neq 0 \qquad\qquad (7.1)$$

where $A_{d,\,t}$ is the level of new infested area that is detected in year *t*, and $A_{s,\,t}$ is the area that is searched in year *t.* Large increases in new detections compared to area search represents a worsening situation for delimitation, so correspond with an increase in *D*, whereas larger areas searched (which are likely to find new infestations) compared to new detections correspond to a decrease in *D*.

Extirpation

Extirpation, *Ex*, refers to the elimination of individual infestations and is derived from the 'monitoring profile' of infestations—the frequency distribution of time since the most recent detection of the weed at infested sites.

In the revised eradograph, progress towards extirpation at time *t* can be represented by the difference between E_{max} and E_{mean}:

$$Ex_t = E_{max} - E_{mean} \qquad\qquad (7.2)$$

where E_{max} is the average time since the final detection after which it may safely be concluded that the population has been extirpated, often equivalent to the seed longevity, and E_{mean} is the mean of the frequency distribution of the time since the most recent detection. As more and more sites are simply monitored and as these sites stay in the 'monitored' stage through time, the value of E_{mean} increases. Extirpation at a site would occur when $E_{mean} = E_{max}$ and for the whole incursion when all sites had achieved this status.