TECHNICAL REPORT

APPROVED: 27 May 2020

doi:10.2903/sp.efsa.2020.EN-1873

Guidelines for statistically sound and risk-based surveys of *Xylella fastidiosa*

European Food Safety Authority (EFSA),
Elena Lázaro, Stephen Parnell, Antonio Vicent Civera, Jan Schans, Martijn Schenk, Gritta Schrader, Jose Cortiñas Abrahantes, Gabriele Zancanaro, Sybren Vos

Abstract

At the request of the European Commission, EFSA prepared specific guidelines for the survey of *Xylella fastidiosa* to guide the surveyor through the design of statistically sound and risk-based surveys, integrating the key biological information. Based on examples, three different survey designs are simulated: detection surveys to substantiate pest freedom, delimiting surveys to determine the boundaries of an infested zone, and buffer zone surveys to monitor a zone ensuring pest detection at a low level of prevalence. The first step of the survey design consists of setting out the aims of the survey, characterising the host plant population and the methods used to identify the pest. All the survey parameters are quantified considering the importance of the related assumptions. The more accurate the information used to select/estimate the survey parameters, the more robust the conclusions of the survey will be. The second step of the survey design consists of the sample-size calculation using the survey parameters as inputs for the statistical tool (RiBESS+). The last step of the survey design is the allocation of the samples in the survey area, the method for which depends on the information available on the target population and risk factors. The robustness of the conclusions of surveys designed using these approaches depends strongly on the survey preparation. The methodology here proposed allows surveys to be compared across time and space, thus contributing to harmonisation of the *X. fastidiosa* surveys in the EU Member States. The extremely flexible approaches allow surveys to be tailored to each specific situation in the Member States, taking into account the host plants, vectors, climate suitability and resources available. The success of a good survey design relies on technical aspects of the survey preparation and on the involvement of the risk managers.

Key words: buffer zone, confidence level, delimiting, design prevalence, detection, epidemiological units, infested zone

Requestor: Directorate-General for Health and Food Safety, European Commission

Question number: EFSA-Q-2020-00235

Correspondence: ALPHA@efsa.europa.eu

Acknowledgements: EFSA wishes to thank the following for the support provided to this scientific output: Makrina Diakaki and Giulia Mattion. EFSA wishes to acknowledge the Member State institutions that contributed to the two workshops held in Parma for the development of the survey approaches for *Xylella fastidiosa*. EFSA also wishes to acknowledge the Netherlands Food and Consumer Product Safety Authority (NVWA) and the Julius Kuehn Institute (JKI) in Germany for the support provided to this scientific output in the context of grant GP/EFSA/ALPHA/2017/02.

Suggested citation: EFSA (European Food Safety Authority), Lázaro E, Parnell S, Vicent Civera A, Schans J, Schenk M, Schrader G, Cortiñas Abrahantes J, Zancanaro G and Vos S, 2020. Guidelines for statistically sound and risk-based surveys of *Xylella fastidiosa*. EFSA supporting publication 2020:EN-1873. 76 pp. doi:10.2903/sp.efsa.2020.EN-1873

ISSN: 2397-8325

Summary

At the request of the European Commission, to support the EU Member States, EFSA prepared specific guidelines for the survey of *Xylella fastidiosa*. This document guides the surveyor through the design of statistically sound and risk-based surveys for *X. fastidiosa*, integrating into the design the key information gathered from the pest survey card for *X. fastidiosa* (EFSA, 2019a).

Three different survey aims are distinguished: *detection surveys* to substantiate pest freedom in an area or country, *delimiting surveys* to determine the boundaries of an infested zone, and *buffer zone surveys* to monitor a zone that serves as a buffer around an infested zone and therefore should ensure pest detection even at low levels of prevalence. The guidelines have been developed using examples to illustrate the design of these three types of survey.

The first step of the survey design is to set the aim of the survey, and to characterise the host plant population as well as the identification method for the pest. It will be necessary to quantify the survey parameters and to consider the importance of the assumptions that are made for each one of them. When setting the design prevalence and the confidence level of the survey, the chosen values should reflect the aim of the survey and the compromise between the resources needed to carry out the survey and the risk that risk managers are willing to accept. Good information on land use in the survey area is needed to determine the size of the target population and its hierarchical structure. The host plant population can then be defined by subdividing it into units that are homogeneous in terms of the epidemiology of *X. fastidiosa.* The use of risk factors will allow the surveys to be better targeted to those areas where the probability of infection is higher. The relative risks can be estimated using expert knowledge or by means of data analysis. The method sensitivity needs to be estimated by combining sampling effectiveness and diagnostic sensitivity, which is particularly challenging for *X. fastidiosa* because the method sensitivity varies depending on the host species and a conservative approach is recommended here. The better the information used to establish the survey parameters, the more robust the conclusions of the survey will be.

In the second step, the sample size is calculated using the survey parameters as input for the statistical tool RiBESS+, which calculates the sample size using a statistically sound and risk-based approach. The mathematical principles behind the tool are fully in line with the recommendations and guidelines provided by the different International Standards for Phytosanitary Measures. In addition, RiBESS+ is routinely used for surveillance activities in the animal health sector. The approach is further tailored to the surveys of *X. fastidiosa* and illustrated using examples.

The final step of the survey design is the allocation of the samples within subdivisions of the target population. Depending on the information available on the target population and risk factors, the allocation of the samples can be proportional to the number of epidemiological units, or to the size of the host plant population or to the number of risk locations in each region of the survey area. If no information is available, the samples could be allocated at random across the entire survey area.

The robustness of the conclusions of the surveys designed using the proposed approach depends strongly on the quality of the design. The proposed methodology allows one to compare surveys across time and space, thus contributing to harmonisation of surveys in the EU MSs.

Considering that the survey obligations are at EU MS level, and that the data required for survey design are available at national or even regional level, the developed approach should be tailored to each specific situation in terms of host plants, vectors, climate and resources. The approach and tools provided for the specific surveys of *X. fastidiosa* are quite flexible and the success of the design procedure relies on the technical aspects of the survey preparation and the involvement of risk managers.

Table of contents

Introduction

At the request of the European Commission, EFSA was asked to support the EU Member States (MSs) in the preparation and planning of the surveys of the EU quarantine pests (EFSA mandate on plant pest surveillance M-2017-0137). In this context, EFSA prepared the *Xylella fastidiosa* survey toolkit that includes: (i) the general survey guidelines (EFSA, in preparation), which describe the context in which the surveys are to be performed (legal, international standards, scientific knowledge), and the basic principles and approaches that are implemented for surveillance of Union quarantine pests; (ii) the Pest survey card on *X. fastidiosa* (EFSA, 2019a), which guides the survey designer through the information needed to prepare the survey; (iii) this document, the specific guidelines for surveys of *X. fastidiosa*, which guide the design of statistically sound and risk-based surveys by integrating the key information from the pest survey card and by processing the information for the estimation and allocation of the sampling effort; and (iv) the statistical software tool RiBESS+[1] for the calculation of the sample size.

The surveillance activities should be planned in three steps:

1. The first step is the *survey preparation*. This is described in the Pest survey card on *X. fastidiosa* (EFSA, 2019a), where the objective is to define and characterise at least qualitatively:

- the target population (extension and structure)
- the epidemiological units
- the risk factors and select the risk sites
- the inspection units
- the detection, sampling and identification methods (time, symptoms, sampling matrix, sampling procedure, lab tests).

2. The second step is the *survey design*. This is described in this document, and it is necessary to quantify each survey parameter (indicating the related assumptions) as inputs for the statistical tool RiBESS+:

- the aim of the survey (design prevalence, confidence level)
- the target population (size)
- the epidemiological units (number of units) and host plant proportion within each unit
- the risk factors (relative risk and the proportion of the host plant population to which each risk factor applies)
- the method sensitivity (sampling effectiveness and diagnostic sensitivity).

The resulting sample size is then allocated to subdivisions of the target population or the target population as a whole.

3. The third step is the *survey implementation* (data collection and reporting). The implementation of the survey in the survey area includes the field inspections, sample collection and testing of samples. For the reporting of the survey activity, the results for each inspection and sample should be collected in a dedicated database. Clearly MSs need to follow good practice in data recording; for example, recording absence data (as well as positive findings), identifying which surveys and inspections the samples originate from, and retaining geographical locations of sampled points. These activities are not addressed in the EFSA toolkit for pest surveys. However, some indications are provided on the underpinning assumptions for confidence level and design prevalence and their influence on the conclusion of the survey. The formulation of the survey conclusion allows the surveillance activities to be compared across the EU MSs, within a country, and from one year to the next.

The purpose of this document is to assist the MSs to plan annual survey activities of quarantine pests using risk-based and statistically sound pest survey approaches, in line with current international

[1] https://shiny-efsa.openanalytics.eu/app/ribess

standards. The main focus is to provide guidance on the survey design. Survey implementation, data collection and the reporting of the survey are not addressed in depth in this document.

The situation varies across MSs and even within a MS in terms of: resources available for surveillance; host plant presence and distribution, and density; vector presence, abundance and phenology; environmental suitability; the potential risk of entry, risk activities and presence of risk locations; available methods for detection, sampling and identification. Therefore, the survey design should be tailored to the specific situation in each MS and the guidelines are explained using fictitious data that have been developed for illustration purposes only, i.e. the scenarios and numbers used in this document are explanatory examples and do not reflect any real situation.

1. Problem formulation

1.1. Legal basis

Xylella fastidiosa is a Union quarantine pest that is known to occur in the EU. It is regulated in the EU as a harmful organism under Plant Health Regulation (EU) 2016/2031[2] on the protective measures against pests of plants, under Annex II part B of the implementing act and it is included in the list of priority pests by Regulation (EU) 2019/1702[3].

Decision (EU) 2015/789[4] specifies the emergency measures taken to prevent the introduction into, and the spread within, the EU. This document lays down the main principles and requirement for conducting a risk-based survey for *X. fastidiosa* in demarcated areas as well as in the rest of the territories. In particular, the Decision specifies the basis for calculating the number of inspections and samples that are required. It has been amended on several occasions based on new scientific developments and on the experience acquired in the EU outbreaks. The Decision is likely to be updated when relevant new information becomes available.

1.2. Key epidemiological issues relevant for surveys

Xylella fastidiosa presents a multitude of challenges for surveillance planning that require a systematic and risk-based approach to allocate the survey effort and a careful selection of detection methods. In particular, infections by *X. fastidiosa* are generally characterised by a long (and variable) asymptomatic period, cryptic expression of symptoms that resemble signs of drought stress, intraspecific diversity and aggressiveness of the bacterium, uncertainty about the host range, and variation in environmental suitability (EFSA PLH Panel, 2019).

1.2.1. Asymptomatic period and symptom expression

Xylella fastidiosa is known to exhibit a long asymptomatic period (i.e. the time from infection of a plant to expression of symptoms). This is a key challenge for surveys, particularly those that are primarily based on visual inspection (Figure 1). After its introduction, *X. fastidiosa* can spread widely before the first host plants display symptoms. Visual inspection surveys may thus not be sufficient for *X. fastidiosa*, depending on the survey objectives. Substantiating an area as free from visually detectable symptoms has limited bearing on whether an area is truly free from infection. Moreover, when surveys are led by visual inspection alone, *X. fastidiosa* outbreaks will in general be detected at a stage in which the prevalence already exceeds the level at which the outbreak can still be eradicated.

[2] Regulation (EU) 2016/2031 of the European Parliament of the Council of 26 October 2016 on protective measures against pests of plants, amending Regulations (EU) 228/2013, (EU) 652/2014 and (EU) 1143/2014 of the European Parliament and of the Council and repealing Council Directives 69/464/EEC, 74/647/EEC, 93/85/EEC, 98/57/EC, 2000/29/EC, 2006/91/EC and 2007/33/EC. OJ L 317, 23.11.2016, pp. 4–104.

[3] Commission Delegated Regulation (EU) 2019/1702 of 1 August 2019 supplementing Regulation (EU) 2016/2031 of the European Parliament and of the Council by establishing the list of priority pests. OJ L 260, 11.10.2019, p. 8–10.

[4] Commission Implementing Decision (EU) 2015/789 of 18 May 2015 as regards measures to prevent the introduction into and the spread within the Union of *Xylella fastidiosa* (Wells et al.) (notified under document C(2015) 3415). OJ L 125 21.5.2015, p. 36.

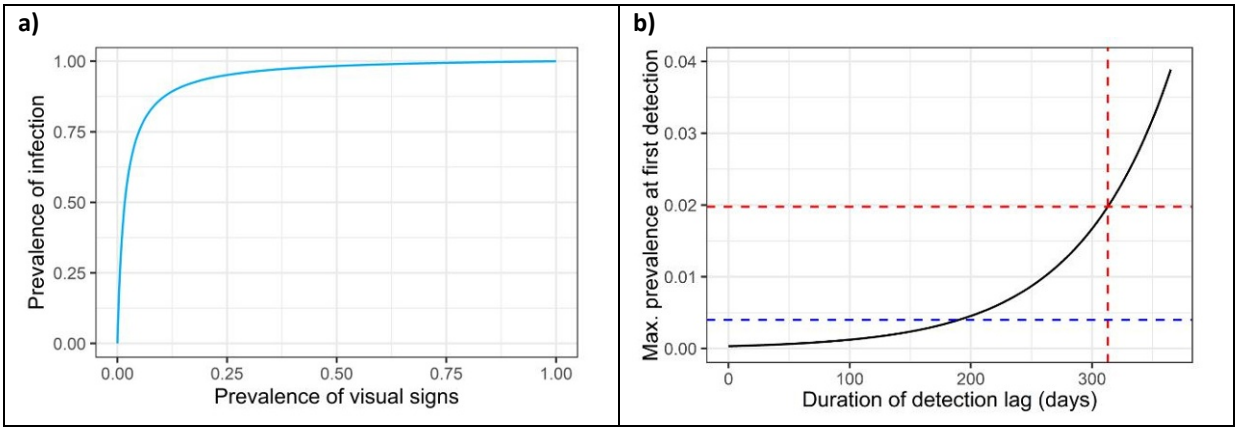


Figure 1: The challenge posed by the long asymptomatic period of *Xylella fastidiosa* for detection surveys and management of the disease. (a) The relationship between the true prevalence of infection with the prevalence of visual symptoms of *X. fastidiosa* following the first discovery of a local outbreak. (b) The prevalence at first discovery assuming a visual survey inspecting 840 trees per day over a 50-day period. Despite this high sampling intensity, the prevalence at first detection is 0.02 (2%) of the population (red dashed line), whereas the estimated eradicable prevalence is much lower at 0.0004 (0.04%) (blue dashed line). Calculations are based on epidemic growth rate data from olive orchards in Apulia (Hornero et al.; 2020), asymptomatic period data for olive trees (EFSA PLH Panel, 2019) and a mathematical framework linking epidemiological parameters and surveillance (Parnell et al., 2017; Mastin et al., 2017, 2019)

In EFSA PLH Panel (2019), based on the literature review reported in the *Xylella* Host Plant Database (EFSA, 2018), studies were identified which quantify the time from infection to the onset of symptoms. The data were categorised by host plant and *X. fastidiosa* subspecies. The analysis of the data revealed a large variation in the length of the asymptomatic period depending on the host–subspecies combination. The asymptomatic period in *Catharanthus roseus*, an ornamental plant, had a median of 30 days, while for olive the median was 390 days (for the *pauca* subspecies) (EFSA PLH Panel, 2019).

This long asymptomatic period is further exacerbated by the cryptic nature of symptoms once expressed. Symptom expression is usually linked to the occlusion of xylem vessels. Hence, symptoms of *X. fastidiosa* tend to resemble those caused by water stress. In some cases, the infection results in rapid death of the host plant (Purcell and Saunders, 1999; Martelli et al., 2016). However, some plant species may not express any symptoms at all, which may in turn depend on the growing conditions (EFSA PLH Panel, 2015, 2018, 2019).

1.2.2. Intraspecific diversity and host range

A challenge for *X. fastidiosa* surveys is the wide range of potential host species that could be infected. There are currently 595 potential host species reported in the scientific literature (EFSA, 2020), 221 of which have been found infected in natural conditions and confirmed by at least two molecular methods. In addition, these numbers of host species relate to the *X. fastidiosa* species as a whole, whereas different subspecies or even sequence types have different host ranges and aggressiveness. However, detection surveys for substantiation of pest freedom operate at the species level, meaning the aim is to establish whether *X. fastidiosa* is present, and not whether a particular subspecies is present. The large number of potential host species deems it impractical to consider surveys for each individual plant species. It is therefore necessary to identify and target species that are more likely to be infected (e.g. because of known vector preferences in a MS or those because they are known to be infected by multiple *X. fastidiosa* subspecies). According to EFSA (2020), the species *Prunus dulcis*, *Prunus avium*, *Polygala myrtifolia*, *Spartium junceum*, *Nerium oleander*, *Rhamnus alaternus* and *Rosmarinus officinalis* have been reported as being susceptible to at least three subspecies of *X. fastidiosa*. Setting up epidemiologically relevant categorisations of species by type of environment or land use categories (e.g. urban, forest, agricultural, other) will also assist the survey design. The

Pest survey card on *X. fastidiosa* (EFSA, 2019a) provides the reasoning that can be used to inform the selection of the host plants for inclusion in the target population.

1.2.3. Vectors of *Xylella fastidiosa*

Vectors play a major role in the epidemiology of *X. fastidiosa* as the natural spread of the bacterium is exclusively through xylem-feeding insects. However, although it provides an aid to the surveillance in place, the vector survey has not been considered in the survey design in this document for the following reasons:

- **Uncertainty on the vector status**: although all xylem-feeding insects are potential vectors of *X. fastidiosa*, there is a high uncertainty on their transmission capabilities. So far, only *Philaenus spumarius* has been proven to transmit the bacterium in natural conditions in the EU. Regarding its transmission capacity, in the EFSA priority pest datasheet (EFSA, 2019b) the probability of *P. spumarius* transmitting the bacterium was estimated with a median of 0.13. *Neophilaenus campestris* and other species have, so far, only been shown to have the capacity to acquire the bacterium in natural conditions, while their ability to transmit the bacterium to a new host plant still needs to be confirmed.
- **Sweeping effectiveness**: for the vector sampling, the proposed method for capturing the insects is by using a sweep net (EFSA, 2019a). The effectiveness of the sweeping activity, i.e. the probability of capturing a vector carrying the bacterium can be estimated by multiplying the probability that the insect has fed on an infected plant by the probability of the insect acquiring the bacterium. The latter has been estimated at around 0.14 (EFSA, 2019b). Therefore, the resulting vector sampling effectiveness is extremely low. The number of vectors that should be captured to survey *X. fastidiosa* would consequently be very high and difficult to achieve.
- **Traceability**: the legislation is based on infected plants. Therefore, a positive finding in a vector would trigger another survey to identify the infected plants.

With the current state of knowledge, it is not possible to base the *X. fastidiosa* status of an area on a vector survey alone. The vector surveys for *X. fastidiosa* could provide additional information on presence and epidemiology and to plant-based surveys but cannot alone be the basis of an *X. fastidiosa*-specific survey. However, this should be reviewed in the light of any new evidence or development of more effective capture methods.

1.2.4. Climatic suitability and uncertainty

Surveys should be carried out in areas of the EU where the bacterium can potentially become established. The climatic suitability for the potential establishment of *X. fastidiosa* is highly variable throughout the EU and also within individual MSs (EFSA PLH Panel, 2019). *Xylella fastidiosa* is known to occur throughout a wide range of climatic zones, including continental climates (e.g. in northern parts of North America) as well as tropical and sub-tropical climates (e.g. in Brazil, California). In large parts of the EU, the climate is not a limiting factor for the pathogen to become established (EFSA PLH Panel, 2019). However, the southern part of the EU has been evaluated as being most at risk. There is, however, considerable uncertainty about recent estimates of climatic suitability, particularly at the subspecies or sequence type level. This is due to lack of data and a potential bias in reported cases. In particular, the lack of reported cases at northern latitudes where symptom expression of *X. fastidiosa* is lacking may bias current climatic suitability estimates. Nonetheless, variability of climatic suitability within a MS may be accounted for when designing the survey.

2. Survey design

The survey parameters are described in the General guidelines for pest surveys (EFSA, in preparation) and their definitions are also provided in the Glossary for this document.

The survey design consists of quantifying each survey parameter as these are the input values which are needed for the statistical tool (i.e. RiBESS+) to estimate the sample size. The survey parameters are:

- **Confidence level and design prevalence.** Both parameters define the aim of the survey.
- **Target population size.** Indicates the size of the host plant population targeted by the survey to which the survey results will apply.
- **Method sensitivity.** This deals with how good the method is to detect the pest when it is truly present. Method sensitivity combines sampling effectiveness and diagnostic sensitivity values.
- **Risk factors.** For a risk-based survey approach each risk factor must be categorised in different levels (risk factor levels) that are quantified by means of their relative risk and the proportion of the target population to which they apply.

In this document the survey design step is illustrated for two case studies:

1. An annual detection survey to substantiate pest freedom.

2. A delimiting survey after a finding of *X. fastidiosa* and the corresponding buffer zone survey once the infested zone has been demarcated.

In order to calculate the number of samples, EFSA has made the RiBESS+ tool available as a free online application to support the surveillance programme managers (available at https://shiny-efsa.openanalytics.eu/app/ribess). Figure 2 shows a screenshot of RiBESS+ for calculating a sample size showing the five above-mentioned input parameters and the calculated output. Additional functionalities are available and described in the RiBESS+ manual[5], but are not used in these guidelines.

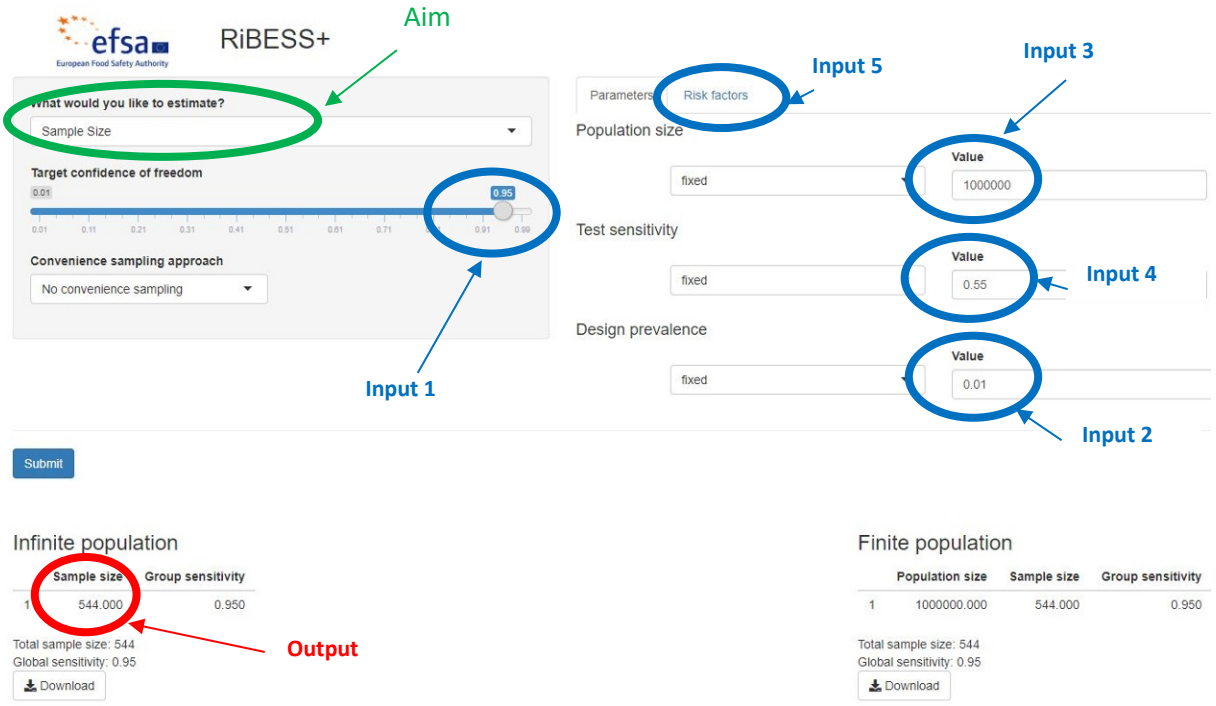


Figure 2: Screenshot of the sample size calculation (Output) using RiBESS+ with a 95% confidence level (Input 1), a 1% design prevalence (Input 2), assuming a population size of 1,000,000 (Input 3), a method sensitivity of 55% (Input 4, combining sampling effectiveness and diagnostic sensitivity) and the risk factors tab (Input 5 – addressed in Section 2.4)

[5] https://zenodo.org/record/2541541/preview/ribess-manual.pdf

2.1. Confidence level and design prevalence

There is not one single approach to designing and implementing a survey and determining the required number of samples. Both the confidence level and design prevalence are pivotal to drawing the survey conclusions. Consequently, the selection of the confidence level and design prevalence values requires finding a compromise between available resources and the level of the risk that the risk managers are willing to accept.

2.1.1. Detection survey

The **confidence level** reflects the level of accuracy (confidence) in the results (conclusions). When it is stated that a territory is 'pest free' with 95% confidence, this means that given the methods and the assumptions taken, the statement is (on average) expected to be correct at least 95% of the time. In general, confidence levels are set at 95%. When setting the confidence level, the risk managers should consider the resources available and the epidemiological situation that might vary in the territory of their MS. For a detection survey for *X. fastidiosa* the confidence level could be set at 95%.

Design prevalence refers to the minimum prevalence that is aimed at detecting whether the pest is circulating in the region surveyed given the sample size, detection methods, and conditions of the survey. This threshold value is usually based on scientific analysis, policy decisions, and risk assessment by all parties involved. Prevalence is the proportion of the population infected and risk managers should note, for different environments and areas, what this means for both the absolute number of infected plants and their spatial distribution. That is, 1% of infection of a small population has different implications for eradication and control than 1% of a large population. Moreover, 1,000 infected trees in a single focus are more readily controlled than 1,000 trees scattered over a large area. In general, the higher the design prevalence, the more likely that an outbreak remains undetected for a prolonged period and will consequently be more widely distributed before it is detected.

For the *X. fastidiosa* detection survey, design prevalence is recommended to be fixed according to two different objectives:

- **Pest absence confirmation**. In any area where *X. fastidiosa* is not yet known to occur, and without neighbouring outbreak areas that present an immediate risk for introduction, one can reasonably assume that the disease is not present. In that situation, the design prevalence acts as a proxy for zero. As per the general guidelines (EFSA, in preparation), there are a number of ways in which the design prevalence can be motivated in this situation (e.g. using epidemiology or a trade-risk model). In the absence of such information, the standard approach in animal health is to use a design prevalence of 1% at the herd level (FAO, 2014). Risk managers are recommended to calculate what this percentage means in terms of the absolute numbers of infected host units to ensure that they find this level of risk acceptable. For example, if the aim of the survey is to have 95% confidence of detecting 1% prevalence, this would mean that if there is no positive finding in an area of 1,000,000 olive trees, one has 95% confidence that the actual number of infected trees is between 0 and 10,000.
- **Pest freedom in an area neighbouring an outbreak**. In areas adjacent to a known outbreak area, the probability of introduction of *X. fastidiosa* will be high and therefore the objective is to ensure that any new infections are detected when still below the extent and prevalence at which eradication is possible.

To determine the prevalence of infected host plants at which eradication is still achievable depends on a range of interacting factors. These include the environmental conditions, vector abundance, host availability and distribution as well as the intensity of the eradication measures that are implemented. EFSA PLH Panel (2019) investigated a range of epidemiological scenarios for the eradication of *X. fastidiosa* outbreaks in olive trees. In the model, some simulated scenarios resulted in successful eradication of the pest when combining vector control, clear cut of host trees, and detecting the pest, three years after an initial introduction event with pest prevalence below 0.4%. Based on these results, a design prevalence of 0.4% would be recommended. However, when setting the design

prevalence, the risk managers should consider that the level of possible eradication also depends on the resources available and the epidemiological situations that might vary in the territory of the MS.

2.1.2. Delimiting survey and buffer zone survey

When a new outbreak has been found, it is necessary to delimit the infested zone where the pest is circulating as quickly as possible (thus using a much lower design prevalence) to avoid further spread of the pest. Subsequently, it is necessary to demarcate an area with a buffer zone around the infested zone to protect the rest of territory from the identified pest.

In this situation, the **confidence level** could be fixed at the same level as was used for the detection survey (e.g. 95%). Based on the values recommended under the two detection survey situations (see Section 2.2 for further details), the **design prevalence** is recommended to be set 10 times lower, at **0.1%** and **0.04%**. Note that the design prevalence of the survey is a compromise between the available resources and an acceptable risk level, and this should be substantiated by evidence.

2.2. Target population

When designing a survey, the target population needs to be clearly identified. EFSA (2018) define the target population as the set of individual plants or commodities or vectors in which the pest under scrutiny can be detected directly (e.g. looking for the pest) or indirectly (e.g. looking for symptoms suggesting the presence of the pest) in a given habitat or survey area. From a practical point of view, one needs to:

i) properly define the inspection units (e.g. **host plants**);

ii) quantify the total number of the inspection units in the survey area (**size**); and

iii) characterise how these inspection units are structured in the landscape and can be included in homogeneous groups according to their epidemiology (e.g. geographical distribution). Each individual homogeneous group of inspection units defines an **epidemiological unit**.

2.2.1. Target population size

The target population size is defined by the total number of inspection units within the survey area. As mentioned in Section 1.2.2, the wide range of potential host species poses a significant challenge for *X. fastidiosa* surveys. Nevertheless, it is necessary to identify the host species under consideration to quantify the target population size. In a detection survey, the survey will probably cover a wide area, and the survey area can be split into four different types of environment depending on the land use to better identify the most relevant host plants:

(i) **Agricultural areas** where host plants are cultivated in fields, orchards or in greenhouses.

(ii) **Urban areas** where host plants are growing as residential and ornamental plants in private gardens and parks or as lane trees.

(iii) **Forest areas** where host plants are growing in managed and natural forests.

(iv) **Other areas** where host plants are growing in natural or semi-natural conditions.

This approach can facilitate the host plant selection for the survey area. The Pest survey card on *X. fastidiosa* (EFSA, 2019a) provides rationales that can be used to select the host plants in areas with similar environmental conditions to the current outbreak areas in the EU. Appendix A shows an example of how the different land use categories can be mapped using the information in the Corine Land Cover database[6].

Within the land use categories, Appendix B shows a host plant ranking (at genus level) according to the probability of being infected by *X. fastidiosa* based on available data from the current outbreak areas in the EU (see Section 2.4 for further details).

[6] https://land.copernicus.eu/pan-european/corine-land-cover

After selecting the host plant species to be targeted by the survey, the size of the target population in the MS has to be estimated based on available information on cultivation, land use, flora inventories, etc.

2.2.2. Epidemiological units

Definition

As defined in the Glossary, an epidemiological unit is a homogeneous area where the interactions between the pest, the host plants, the abiotic and biotic factors and conditions would result in a similar epidemiology, should the pest be present. The epidemiological units are subdivisions of the target population according to an epidemiological homogeneity criterion and reflect the structure of the target population in a given geographical area. They are the units of interest for which the sample size is estimated. This could be achieved by calculating the overall sample size and then proportionally allocating them to each subdivision of the target population. For a statistically based survey it is therefore essential to define the epidemiological units and clearly indicate the underlying assumptions.

Here, we illustrate two extreme cases to define the epidemiological unit:

(i) The whole survey area is considered as one independent epidemiological unit. This homogeneity assumption would rarely be fulfilled as the epidemiology usually varies across larger areas in terms of ecology (habitats, environmental suitability, timing of life stages in the year, crops, host plants, vector abundance, etc.), exposure (pathways and entry points, flora, etc.), geographical and topographical characteristics.

(ii) Each hectare (for the whole survey area or for each land use category) that contains at least one host plant is considered an independent epidemiological unit. This case applies when there is little available information on the homogeneity. Appendix C expands this case by showing how the survey design can be developed in a demarcated area by first estimating the sample size needed within the single hectare and then by estimating the number of hectares that need to be inspected. The resulting sample size should be calculated by multiplying the number of hectares that need to be surveyed by the sample size calculated within the single hectare.

To optimise the survey efforts in terms of the number of samples that represent the host population, it is essential to gather as much information as possible on the homogeneity of the territory and to choose an epidemiological unit size in which the homogeneity assumption is realistic and acceptable.

For the annual detection survey in a MS, as an example of an intermediate situation, the NUTS regions (Nomenclature of Territorial Units for Statistics) could be considered for defining the epidemiological units. The NUTS classification is a hierarchical system for subdividing the economic territory of the EU. Details on all NUTS regions in the EU are available from Eurostat (2018). When using the NUTS region classification in the survey design the epidemiology of *X. fastidiosa* is assumed to be the same within each area. Obviously, the bigger these regions are the more difficult it will be to fulfil the assumption of homogeneity of the survey area.

For the delimiting and buffer zone surveys the homogeneity criteria could be set by considering the different land use categories (Sections 2.2.1 and 3 for further details) as independent epidemiological units given that the focus of the survey is at a local level.

Sample size within an epidemiological unit

It is possible to calculate the sample size for an epidemiological unit given a specific confidence level, design prevalence, target population size and method sensitivity.

Figure 3 shows how the estimated sample size for an unknown (statistically infinite – binomial distribution) population size compares to a known population size (finite population – hypergeometric distribution), given a confidence level of 95%, a design prevalence of 1% and a method sensitivity of 80%. At a certain population size, the host population can be considered as infinite from a statistical point of view. In the illustrated case, both curves converge around 15,000 host plants, which would require about 370 samples. Above 15,000 plants, only very few additional samples are needed to

achieve the same confidence of 95% and design prevalence of 1% (15,000 – 370 samples; 20,000 – 371 samples; 60,000 – 373 samples). As a consequence, when little information is available about the host population size within an epidemiological unit, it is still possible to determine the sample size without knowing the exact number of host plants.

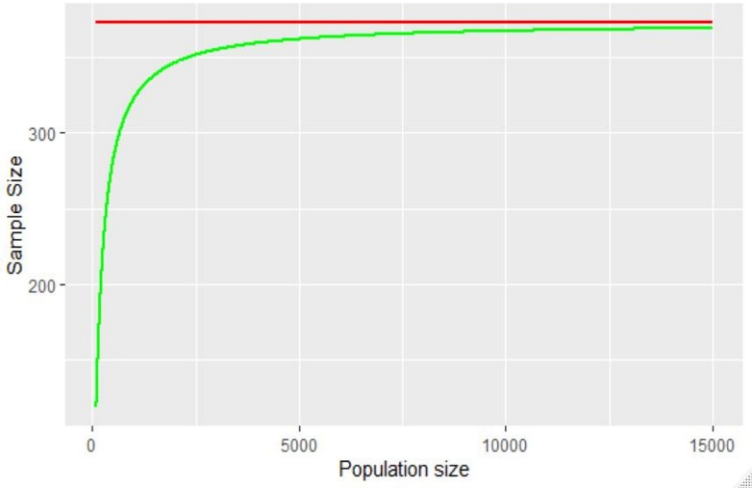


Figure 3: Within an epidemiological unit, for a given method sensitivity of 80%, a confidence level of 95% and a design prevalence of 1%, the sample size follows a hypergeometric distribution (green) for finite population sizes, and a binomial distribution (red) for infinite populations

Table 1 shows the estimates of sample size under different confidence levels and design prevalence values for an epidemiological unit with an infinite population of host plants and a method sensitivity of 55% (see Section 2.3).

Table 1: Sample size estimates using RiBESS+ with different confidence level and design prevalence values for an epidemiological unit with an infinite population of host plants assuming a method sensitivity of 55%

Confidence levels	Design prevalence			
	Demarcated area surveys		Detection surveys	
	0.04%	**0.1%**	**0.4%**	**1%**
99%	20,931	8,371	2,091	835
95%	13,616	5,446	1,361	544
90%	10,446	4,186	1,046	418
85%	8,623	3,449	862	344
80%	7,315	2,926	731	292
53%	3,432	1,373	343	137

The red values are used in the various case studies, for annual detection survey as well as for delimiting and buffer zone surveys.

2.2.3. Hierarchical structure of the target population

As described in the previous sections, the target population is structured at different levels, i.e. host plants in the MS; host plants for each land use category; host plants within the epidemiological units; and inspection units. Figure 4 illustrates these different hierarchical levels which need to be defined during the survey design.

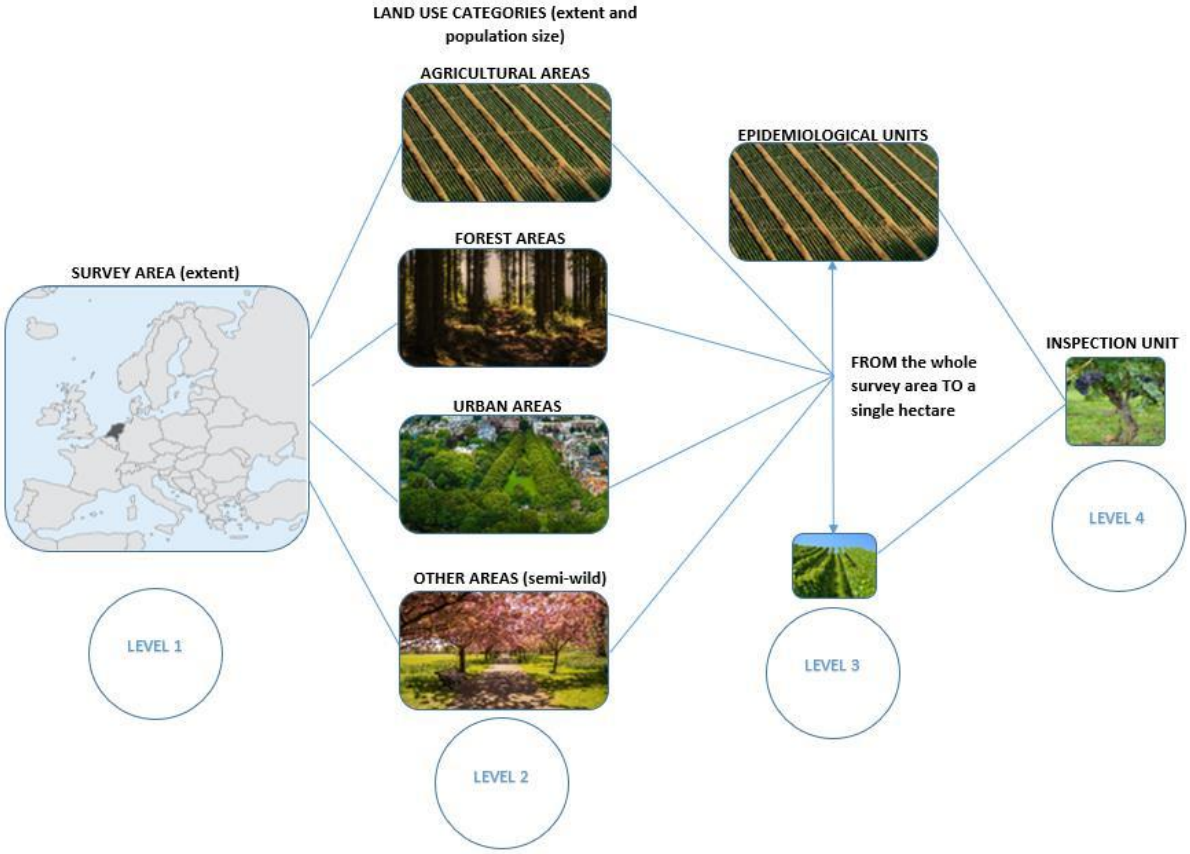


Figure 4: Hierarchical structure of the target population: survey area (level 1), land use categories (level 2), epidemiological units (level 3) and inspection units (level 4)

2.3. Method sensitivity

Once the target population is clearly identified, it is necessary to identify the procedures that the inspectors and technicians will follow for:

(i) the field inspection and the visual examination of the host plants (inspection units);

(ii) taking samples; and

(iii) the identification of the pest during laboratory analysis of the samples.

Together, these procedures constitute the overall method of detection and identification for which it is necessary to estimate the sensitivity. The **method sensitivity** is defined as the probability that a truly positive host tests positive. It has two components, the **sampling effectiveness** (i.e. the probability of selecting infected parts from an infected plant, as it is assumed that the diagnostic method cannot be applied to the whole host and a selection of the material to be analysed would be necessary) and the **diagnostic sensitivity** (characterised by the laboratory test used in the identification process).

The **sampling effectiveness** depends on the ability of the inspector to successfully choose the infected plant parts from a host plant. It is directly linked to the sampling procedure itself and on the expertise of the inspectors to recognise any symptoms of the pest. Furthermore, symptom expressions are going to be dependent on other factors, such as the weather conditions and the

physiological stage of the host plant. The collection of the relevant plant parts in the field must follow the procedures recommended by the National Plant Protection Organisation (NPPO). Each sample is georeferenced and consists of a number of plant parts as recommended by the EPPO and IPPC diagnostic protocols for the different host plant species (FAO, 2018; EPPO, 2019). For the case studies in this document an average **sampling effectiveness of 0.70** has been applied.

The **diagnostic sensitivity** is the probability that a sample tests positive when the sample is truly positive. This parameter is provided by the laboratory performing the tests. In the EPPO diagnostic protocol this value is provided for different tests and for different host plants and sampling matrices. For example, when using a specific PCR method on olive samples this value is estimated at 0.67, while for *Polygala* samples it is set at 0.90 when using the same PCR method. For the case studies in this document a **diagnostic sensitivity of 0.78** has been applied to reflect a hypothetical host that would be in between these matrices.

The overall method sensitivity for plant samples from the field to the laboratory can then be calculated:

Method sensitivity = sampling effectiveness × diagnostic sensitivity = 0.70 × 0.78 = **0.55**,

thus 0.55 can be considered as a reference value.

2.4. Risk factors

Consideration of risk factors in the survey design allows the survey efforts to be enforced in those areas where the probabilities of finding the pest are highest. The risk factors that are relevant for surveillance are those that have more than one level of risk for the target population. To be able to use a risk factor in the survey design, it is necessary to characterise both the relative risk and the proportion of the overall plant population to which it applies. In some cases, MSs may have access to detailed and well-parameterised epidemiological spread models (e.g. White et al., 2017; Martinetti and Soubeyrand, 2019) or climate suitability models (e.g. EFSA PLH Panel, 2019) that can be used to determine risk factors. This will depend on data availability and the epidemiological situation. In situations where less information is available, different risk factors can still be distinguished. For example, a risk factor could be defined based on the distance from risk locations. The probability of infection of a host plant species when exposed to the bacterium could be another risk factor.

2.4.1. Distance from the risk locations

As described in EFSA (2019a), it is necessary to first identify the risk activities that could contribute to the introduction or spread of *X. fastidiosa* before the identification of the risk areas. These activities should be connected to specific locations that are then called 'risk locations'. In consideration of the spread capacity of the pest and the availability of host plants, risk areas around these locations can then be defined. The type of locations corresponding to the risk activities related to *X. fastidiosa* are summarised in Table 2. Different relative risk levels may be assigned to the locations depending on traits such as handling plant material from infested areas or being found in an area of dense production involving host plant species.

Table 2: Risk activities and corresponding risk locations relevant for surveillance of *Xylella fastidiosa* in all EU Member States

Risk activity	Risk locations
Production, storage and handling of host plants for planting	• Nurseries and garden centres cultivating or storing ornamental plants, crop plants or saplings for planting
Transport of propagating material	• Stops along main roads and railways (e.g. truck parking lots) for routes connected to infected areas • Airports and harbours with movement from infected countries or areas
Tourism	• Host crops, gardens, parks in the vicinity of tourist sites
Historical findings	• Eradicated outbreak areas, and locations where positive inspections were performed

Risk area

The risk areas can be defined as areas contiguous to the risk locations. The definition of risk areas around a certain risk location takes the spread capacity of the vector and the availability of host plants into consideration. Based on the indicative distance values for the yearly spread of disease, risk areas can be defined around these locations.

For a detection survey i.e. when no positive finding has yet been reported, the objective is to substantiate pest freedom or to detect the bacterium. Assuming that in a suitable environment a host plant remains persistently infected and that competent vectors are present, the radius from the risk location where the pest is most likely to be found should be approximately 150 m (EFSA, 2019a based on EFSA PLH Panel, 2019). Figure 5 illustrates how to extend a risk area in the proximity of a risk location (i.e. a nursery). Once the risk area has been established, the proportion of the host population within the risk area should be determined and the relative risk of the area should be estimated compared with a baseline area. For instance, the risk of infection of the host plants within this risk area could be considered to be twice as high as for the host plants in other fields.

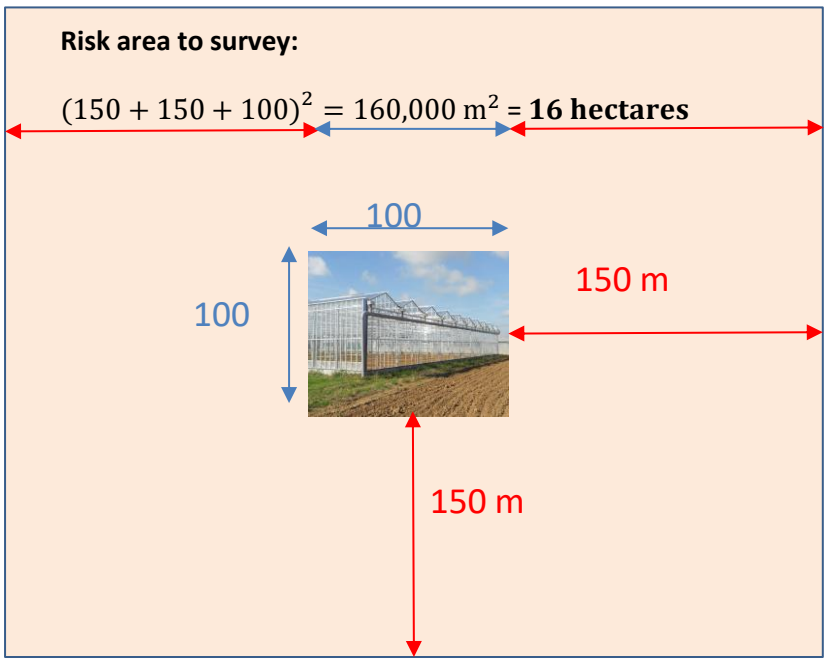


Figure 5: Illustration of how to establish a risk area around a risk location for *Xylella fastidiosa* in a detection survey

For a buffer zone survey, once the potential infested area has been properly delimited (see Section 4 for further details), a buffer zone with a width of 10 km (the median of the long-distance spread model presented in EFSA PLH Panel (2019)) could be demarcated. It is recommended that a high-risk band of 400 m from the boundaries of the infested zone is established as part of the buffer zone. The 400 m corresponds to the upper bound of the short-distance disease dispersal from the model fitted to the Apulian data (EFSA, 2019a based on EFSA PLH Panel, 2019). All targeted host plants within this band can be considered to have twice the risk of being infected compared with host plants in the baseline area. Figure 6 shows how different risk areas can be distinguished in a buffer zone.

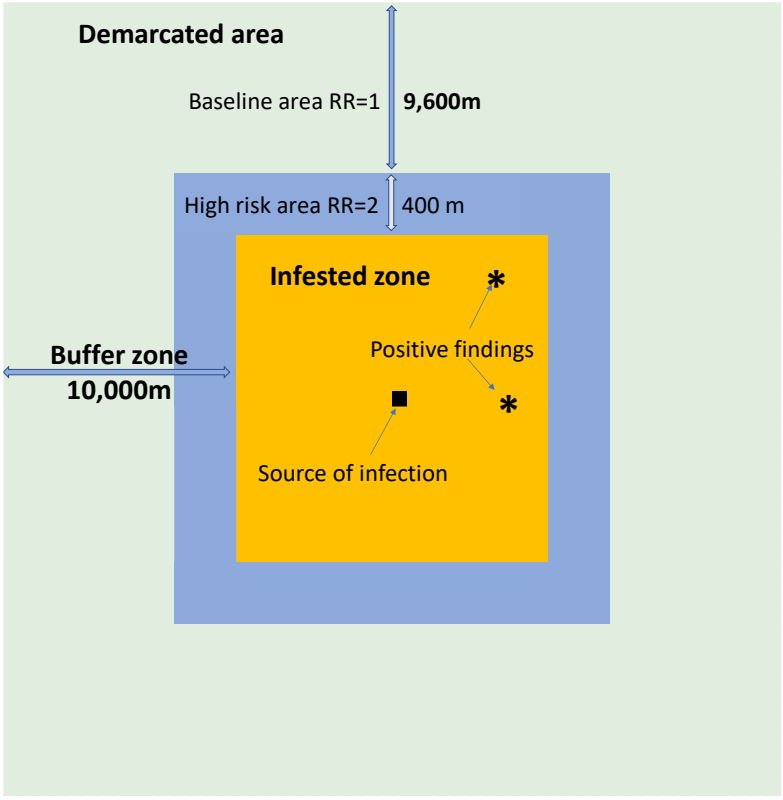


Figure 6: Establishing a risk area under a buffer zone survey

2.4.2. Probability of infection of the host plant species

The wide host range of *X. fastidiosa* complicates the survey design for this pest. Diagnostic methods do not perform the same way on all host plant species and matrices. At the same time, some species may be more susceptible to *X. fastidiosa* than others or may be detected more easily when infected (for instance, hosts that present severe pest-specific symptoms).

Based on the data collected so far in the current EU outbreak zones, the probability of infection of host plants has been estimated at genus level (see Appendix B for further details). Table 3 lists the estimated probability of infection by *X. fastidiosa* of the host plant genera in decreasing order.

Table 3: Estimated probability of infection of the host plant genera being infected by *Xylella fastidiosa* according to the information available from the current EU *Xylella* outbreaks

Genera	Probability of infection
Polygala	0.551
Helichrysum	0.511
Euryops	0.471
Calicotome	0.452
Genista	0.315
Spartium	0.161
Lavandula	0.152
Cistus	0.126
Prunus	0.093
Olea	0.076
Vitis	**0.057**

When the genera in Table 3 are present in the survey area, this information can be used as a guide in the selection of host plants to be surveyed and to integrate this information into the risk-based survey design. To manage this information as a risk factor, each genus must be characterised by its **relative risk** and its **proportion** within the target population. Relative risk estimates can be computed as a ratio of the probability of infection of genus A versus genus B (reference genus considered to be the one present in the area with the lowest probability). Thus, the relative risk can vary depending on the choice of the reference genus and it can be adapted in each survey design according to the most representative genus observed in each landscape. For instance, based on Table 3, when selecting host plants for *X. fastidiosa* surveys, the genera that best characterise the agricultural areas are *Prunus*, *Olea* and *Vitis*, whereas for forest areas these are *Calicotome*, *Lavandula*, *Prunus* and *Olea.* Considering *Vitis* as a reference genus for the agricultural areas (given that it is the genus with the lowest probability of infection from those considered representative in agricultural areas) and *Olea* for the forest areas, Table 4 shows the respective relative risk estimates.

Table 4: Estimated probability of infection and relative risk of the most representative genera for survey in agricultural and forest areas (for illustrative purposes)

Land use category	Genera	Probability of infection	Relative risk
Agricultural areas	*Prunus*	0.093	1.631
	Olea	0.076	1.333
	Vitis	**0.057**	**1.000**
Forest areas	*Polygala*	0.551	7.250
	Calicotome	0.452	5.947
	Lavandula	0.152	3.625
	Prunus	0.093	1.224
	Olea	**0.076**	**1.000**

2.5. Practical considerations

When designing a survey, the underpinning assumptions related, in particular, to the homogeneity of the epidemiological units and the aim of the survey in terms of confidence level and associated design prevalence, need to be clearly formulated and accepted by the risk managers. The assumptions taken will strongly impact on the quantification of the parameters that will determine the sample size for the survey and thus have a strong influence on the reliability of the survey conclusions.

Below, based on the above-mentioned information about target population and risk factor quantification, we provide some general advice on how to combine design prevalence and confidence level threshold values in a practical way to estimate sample size and to interpret survey results.

Design prevalence

Table 5 shows different reference values of the design prevalence according to the survey aim and the different land use categories into which target population can be split. For a **detection survey**, two strategies are considered:

- Set a design prevalence of 1% for all land use categories (row 1).
- Set a design prevalence of 0.4% for agricultural areas and 1% for the rest (row 2). This strategy is based on the assumption that the census availability and accessibility in agricultural areas make surveys more feasible here, so a lower design prevalence could be achieved. However, depending on survey territory characteristics, the main interest could be to focus the survey on other land use categories.

For **delimiting and buffer zone surveys**, the design prevalence can be set by combining the reference threshold values of 0.1% and 0.04% (see Section 2.1.2) in the different land use category areas. As displayed in Table 5 (rows 3 and 4), one possibility is to set the most restrictive value for the agricultural areas, assuming that there is a need to protect this land use category more than the others. However, other strategies might better describe the situation of the disease in a specific territory or MS.

In the logical sequence of surveys, in areas where the pest is not known to occur, detection surveys are first conducted to confirm the pest-free status. Only when the first infection is found are delimiting and buffer zone surveys conducted. The gradient of design prevalence shown as an example in Table 5 follows this logical sequence of surveys.

Table 5: Examples of the risk manager's choice of design prevalence for the different types of survey for *Xylella fastidiosa*

Design prevalence for *Xylella* surveys	Agricultural areas	Urban areas	Forests	Other areas
Annual detection survey	1%	1%	1%	1%
	0.4%	1%	1%	1%
Delimiting surveys	0.04%	0.1%	0.1%	0.1%
Buffer zone surveys	0.04%	0.1%	0.1%	0.1%

Confidence level

- **Case 1:** if the available information allows, for each of these different land use categories, the relative risks and the proportion of the target population affected by these relative risks to be estimated, then it is possible to calculate the sample size for each area. In addition, it allows a general conclusion on pest freedom for the entire target population to be drawn with a given confidence level.
- **Case 2:** if the available information is not sufficient to properly define the land use categories in terms of relative risk and proportion of the target population, then it is still possible to

reach a general conclusion on pest freedom for the entire target population. If the overall confidence level of the survey is set at, for example, 95% the confidence level to achieve in the survey of each one of these land use categories can be calculated (Cannon, 2002) using the following formula:

$$CL = 1 - \prod_{i=1}^{n} (1 - CLi)$$

where *CL* is the overall confidence level of the survey, *CLi* is the confidence level of the survey of the land use category *i* and *n* is the number of different land use categories.

In this specific example, *n* = 4 and *CL* = 0.95 and the confidence level to achieve for each land use category within this survey is obtained:

$$0.95 = 1 - (1 - CLi)^4$$

$$CLi = 1 - \sqrt[4]{0.05} = 0.53$$

This means that if a 53% confidence level is achieved in the survey of each land use category then the pest freedom conclusion for the entire target population can be given with a 95% confidence level.

Additionally, if the goal is to reach a conclusion for each of the land use categories at a confidence level of 95%, then by applying the formula above, the overall survey conclusion can be given with 99.999% confidence. Both settings are developed in the examples in Section 3 where for each one of the four land use categories, these two confidence levels of 95% and 53% are used and the resulting sample sizes compared.

Combinations of different levels of design prevalence and formulation of conclusions

In order to provide an overall conclusion combining different design prevalence when the population is subdivided into different land use categories, the formula given above for Case 2 can be used. For this it is needed to estimate the confidence level achieved for each land use category corresponding to a single value of the design prevalence. If we consider the design prevalence of 0.4% for agricultural areas as the one selected, then the confidence for the other three land use categories is estimated by considering the number of samples collected in each land use category and combining it with the confidence achieved for the agricultural area in the following way:

$$OCL = 1 - (1 - CL_{OLU})^3 \cdot (1 - CL_A)$$

where *OCL* is the overall confidence level, CL_A is the confidence level for an agricultural area, and CL_{OLU} is the confidence level for other land use categories (urban, forest, wild/semi-wild).

Figure 7 illustrates how to estimate the confidence achieved for a specific design prevalence and number of samples taken, using the RiBESS+ tool. It shows that when 137 trees are sampled this corresponds to 53% confidence to detect 1% design prevalence and is equivalent to 26% confidence to detect 0.4% design prevalence.

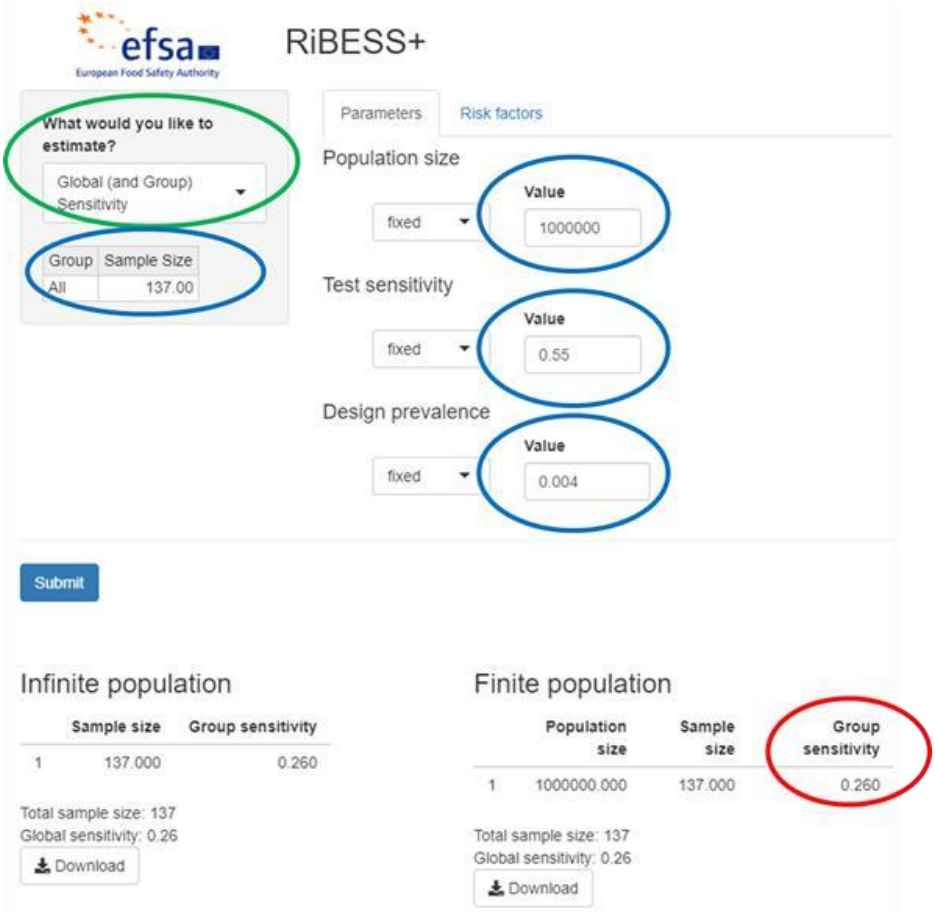


Figure 7: Screenshot of RiBESS+ illustrating the estimation of the confidence achieved for a given population (1,000,000), method sensitivity (0.55), design prevalence (0.004) and sample size (137). Global (and group) sensitivity is the achieved confidence (0.26). The green circle is the chosen functionality, the blue circles are the input values of the survey parameters and the red circle is the estimated output

3. Detection survey for pest freedom substantiation

3.1. Example using NUTS regions as epidemiological units

For the annual detection survey in a MS, the NUTS regions are considered as the epidemiological units (see Section 2.3) in the examples below. The Netherlands (NL) is used as an example with two simulations performed using different definitions of epidemiological units based on the NUTS regions classification. For each option, the sample size needed to perform a statistically based annual detection survey has been calculated. The numbers reflect a hypothetical situation for illustration purposes only and do not reflect the real situation.

3.1.1. NUTS regions in the Netherlands

Figure 8 and Table 6 describe the NUTS regions for the Netherlands and Table 7 provides a summary of their numbers and extent.

Figure 8: The NUTS 3 regions of the Netherlands (Eurostat, 2018)

Table 6: NUTS regions (level 1, 2 and 3) in the Netherlands (Eurostat, 2018)

Code	NUTS 1	NUTS 2	NUTS 3
NL1	**NOORD-NEDERLAND**		
NL11		Groningen	
NL111			Oost-Groningen
NL112			Delfzijl en omgeving
NL113			Overig Groningen
NL12		Friesland	
NL124			Noord-Friesland
NL125			Zuidwest-Friesland
NL126			Zuidoost-Friesland
NL13		Drenthe	
NL131			Noord-Drenthe
NL132			Zuidoost-Drenthe
NL133			Zuidwest-Drenthe
NL2	**OOST-NEDERLAND**		
NL21		Overijssel	
NL211			Noord-Overijssel
NL212			Zuidwest-Overijssel
NL213			Twente
NL22		Gelderland	
NL221			Veluwe
NL224			Zuidwest-Gelderland
NL225			Achterhoek
NL226			Arnhem/Nijmegen
NL23		Flevoland	
NL230			Flevoland

Code	NUTS 1	NUTS 2	NUTS 3
NL3	**WEST-NEDERLAND**		
NL31		Utrecht	
NL310			Utrecht
NL32		Noord-Holland	
NL321			Kop van Noord-Holland
NL323			IJmond
NL324			Agglomeratie Haarlem
NL325			Zaanstreek
NL327			Het Gooi en Vechstreek
NL328			Alkmaar en omgeving
NL329			Groot-Amsterdam
NL33		Zuid-Holland	
NL332			Agglomeratie 's-Gravenhage
NL333			Delft en Westland
NL337			Agglomeratie Leiden en Bollenstreek
NL33A			Zuidoost-Zuid-Holland
NL33B			Oost-Zuid-Holland
NL33C			Groot-Rijnmond
NL34		Zeeland	
NL341			Zeeuwsch-Vlaanderen
NL342			Overig Zeeland
NL4	**ZUID-NEDERLAND**		
NL41		Noord-Brabant	
NL411			West-Noord-Brabant
NL412			Midden-Noord-Brabant
NL413			Noordoost-Noord-Brabant
NL414			Zuidoost-Noord-Brabant
NL42		Limburg	
NL421			Noord-Limburg
NL422			Midden-Limburg
NL423			Zuid-Limburg

Table 7: Numerical description (average, minimum and maximum area) of the NUTS level (1, 2 and 3) regions in the Netherlands (Eurostat, online)

	NUTS level 1 regions			NUTS level 2 regions			NUTS level 3 regions		
	Average	**Min.**	**Max.**	**Average**	**Min.**	**Max.**	**Average**	**Min.**	**Max.**
Area (ha)	1,038,500	729,000	1,189,200	346,200	144,900	574,900	103,900	12,800	343,700
Number of NUTS regions		4			12			40	

3.1.2. Two options using NUTS regions as epidemiological units

Currently the pest status declared by the Dutch NPPO is that *X. fastidiosa* is 'absent, confirmed by survey, intercepted only (2018-04)'. In this case study, the survey is designed to substantiate the freedom from *X. fastidiosa* in the Netherlands. All the information provided above is used to design an annual detection survey for *X. fastidiosa* in the Netherlands with two different definitions of epidemiological units.

Option 1

- NUTS 2 regions are considered as the epidemiological units for surveying the agricultural areas, where the risk of introduction of the pest is assessed as being higher than in the other areas. The assumption is that in the agricultural areas of each NUTS 2 region in the Netherlands the epidemiology of *X. fastidiosa* is similar.
- NUTS 1 regions are considered as the epidemiological units for surveying the urban areas, forest areas and the other areas such as semi-wild or wild areas. The assumption is that in the urban, forest and other (wild or semi-wild) areas of each NUTS 1 region in the Netherlands the epidemiology of *X. fastidiosa* is similar.

Option 2

- NUTS 3 regions are considered as the epidemiological units for surveying the agricultural areas, where the risk of introduction of the pest is assumed to be higher than in the other areas. The assumption is that in the agricultural areas of each NUTS 3 region in the Netherlands, the epidemiology of *X. fastidiosa* is similar.
- NUTS 2 regions are considered as the epidemiological units for surveying the urban areas, forest areas and the other areas such as semi-wild or wild areas. The assumption is that in the urban, forest and other (wild or semi-wild) areas of each NUTS 2 region in the Netherlands, the epidemiology of *X. fastidiosa* is similar.

3.2. Summary table of the input values for the survey parameters

Table 8 summarises the input parameters for calculating the sample sizes using RiBESS+ for the different options describe above. It is to be noted that the numbers in Table 8 are presented for illustration purposes only and are not real data from the Dutch NPPO.

The preparation of this table finalises the initial phase of the survey design as all the parameters required for calculating the sample size have been quantified and it is considered that the related assumptions have been formulated and accepted by the risk managers.

Table 8: Input values of the survey parameters for the Dutch case study for the design of an annual detection survey with two different definitions of epidemiological units

	Item		Urban areas	Agricultural areas	Forest areas	Other areas (e.g. semi-wild)
Aim of the survey	**Confidence levels**	**Overall 95%**	0.53	0.53	0.53	0.53
		Overall 99.99%	0.95	0.95	0.95	0.95
	Design prevalence		0.01	0.01	0.01	0.01
			0.01	0.004	0.01	0.01
Target population in the member State	***X. fastidiosa* host plants**		*Lavandula* sp.	*Prunus* sp.	*Quercus* sp.	*Genista* sp.
	Number of host plants		4 million	20 million	50 million	50 million
Epidemiological units	**Option 1**		Host plants in the urban areas of a NUTS 1 region	Host plants in the agricultural areas of a NUTS 2 region	Host plants in the forests of a NUTS 1 region	Host plants in the other areas of a NUTS 1 region
			4 NUTS 1 regions	12 NUTS 2 regions	4 NUTS 1 regions	4 NUTS 1 regions
	Option 2		Host plants in the urban areas of a NUTS 2 region	Host plants in the agricultural areas of a NUTS 3 region	Host plants in the forests of a NUTS 2 region	Host plants in the other areas of a NUTS 2 region
			12 NUTS 2 regions	40 NUTS 3 regions	12 NUTS 2 regions	12 NUTS 2 regions
Detection method	Inspection unit		Single host plant			
	Detection and identification		Visual examination of symptoms, sampling host plants following the NPPO procedure, testing the samples using RT PCR			
	Method sensitivity		0.55 = 0.70 × 0.78			

3.3. Sample size and sample allocation

The steps in the survey design are:

(i) Calculating the sample size using the RiBESS+ tool. The statistical concepts behind this tool are not discussed here. Details on the freedom from disease approaches are provided in the general guidelines for a pest survey (EFSA, in preparation) and can be found in the literature (Cameron, 2012; FAO, 2014; Cannon, 2002).

(ii) Allocating the samples within the MS territory.

3.3.1. Sample size calculation

Table 9 shows the sample size (number of inspection units) calculated for each one of the different land use categories for the Dutch detection survey with a design prevalence of 1% to achieve an overall (i) 99.999% or (ii) 95% confidence level.

Table 9: Number of inspection units (single host plants) to sample in the different land use areas, considering an infinite host plant population, a fixed method sensitivity of 0.55, to achieve a survey sample with an overall confidence level of 99.999*% (individual areas confidence level of 95%) or 95*% (individual areas confidence level of 53%) and a design prevalence of 1%

Land use	Design prevalence (%)	Confidence level (%)	Sample size	Confidence level (%)	Sample size
Urban areas	1	**95**	544	**53**	137
Forest	1	**95**	544	**53**	137
Other areas (semi-wild)	1	**95**	544	**53**	137
Agriculture	1	**95**	544	**53**	137
Total	**1**	**99.99***	**2,176**	**95***	**548**

*Overall confidence level of the survey calculated using the formula from Section 2.5.

Table 10 shows the sample size (number of inspection units) calculated for each one of the different land use categories for the Dutch detection survey to achieve a survey sample with two overall levels of confidence (99.999% and 95%) and two levels of design prevalence (0.4% for the agricultural areas and 1% for the other areas).

Table 10: Number of inspection units (single host plants) to sample in the different land use categories, considering an infinite host plant population and a fixed method sensitivity of 0.55, to achieve a survey sample with an overall confidence level of 99.999*% (individual areas confidence level of 95%) or 95*% (individual areas confidence level of 53%) and a design prevalence of 1% for urban, forest and other areas and 0.4% for agricultural areas

Land use	Design prevalence (%)	Confidence level (%)	Sample size	Design prevalence (%)	Confidence level (%)	Sample size
Urban areas	1	**95**	544	1	**53**	137
Forest	1	**95**	544	1	**53**	137
Other areas (semi-wild)	1	**95**	544	1	**53**	137
Agriculture	0.4	**95**	1361	0.4	**53**	343
Total	**0.4**	**99.86***	**2,993**	**1**	**98.43***	**754**

*Overall confidence level of the survey calculated using the formulas from Section 2.5.

3.3.2. Proportional allocation of the inspection units to the epidemiological units

The number of inspection units (the 'sampling effort') needs to be distributed within the Member State. This can be done, for example, by proportionally allocating the samples equally in each epidemiological unit as shown in Tables 11 and 12 for Option 1 and Tables 13 and 14 for Option 2.

3.3.2.1. Option 1: NUTS 2 regions for agricultural areas, NUTS 1 regions for urban areas, forests and other areas

Table 11: Proportional allocation of inspection units to each epidemiological unit for each land use category for Option 1 (NUTS 2 regions for agricultural areas, NUTS 1 regions for urban, forest and other areas) and for a fixed design prevalence scenario in the entire territory (1%)

Land use	Design prevalence (%)	NUTS level	Number of epidemiological units	Samples 99.999% confidence level	Allocation of samples	Samples 95% confidence level	Allocation of samples
Urban areas	1	1	4	544	544/4 = 136	137	137/4 = 35
Forest	1	1	4	544	544/4 = 136	137	137/4 = 35
Other areas (semi-wild)	1	1	4	544	544/4 = 136	137	137/4 = 35
Agricultural	1	2	12	544	544/12 = 46	137	137/12 = 12
Total				**2,176**	**2,184**	**548**	**564**

Conclusion derived from Table 11:

Assuming that:

(i) in the four NUTS 1 regions of the Netherlands, the epidemiology of *X. fastidiosa* is similar in all the urban areas, in all the forests and in all the other areas; and

(ii) in the 12 NUTS 2 regions of the Netherlands, the epidemiology of *X. fastidiosa* is similar in all the agricultural areas,

after implementing this survey, should all the samples test negative, it could be concluded that:

(i) with an overall 99.999% confidence for the territory of the Netherlands (or with a 95% confidence for each land use category), if *X. fastidiosa* is present, the number of infected plants is below 1%; and

(ii) with an overall 95% confidence for the territory of the Netherlands (or with a 53% confidence for each land use category), if *X. fastidiosa* is present, the number of infected plants is below 1%.

Table 12: Proportional allocation of samples in each epidemiological unit for Option 1 (NUTS 2 regions for agricultural areas, NUTS 1 regions for urban areas, forests and other areas) and for a variable design prevalence scenario depending on the land use (1% for urban, forest and other areas and 0.4% for agricultural areas)

Land use	Design prevalence (%)	NUTS level (Option 1)	Number of epidemiological units	Samples 99.999% confidence level	Allocation of samples	Samples 95% confidence level	Allocation of samples
Urban areas	1	1	4	544	544/4 = 136	137	137/4 = 35
Forest	1	1	4	544	544/4 = 136	137	137/4 = 35
Other areas (semi-wild)	1	1	4	544	544/4 = 136	137	137/4 = 35
Agricultural	0.4	2	12	1,361	1,361/12 = 114	343	343/12 = 29
Total				**2,993**	**3,000**	**754**	**768**

Conclusion derived from Proportional allocation of samples in each epidemiological unit for Option 1 (NUTS 2 regions for agricultural areas, NUTS 1 regions for urban areas, forests and other areas) and for a variable design prevalence scenario depending on the land use (1% for urban, forest and other areas and 0.4% for agricultural areas)

Assuming that:

(i) in the four NUTS 1 regions of the Netherlands, the epidemiology of *X. fastidiosa* is similar in all the urban areas, in all the forests and in all the other areas; and

(ii) in the 12 NUTS 2 regions of the Netherlands, the epidemiology of *X. fastidiosa* is similar in all the agricultural areas,

after implementing this survey, should all the samples test negative, it could be concluded that:

- with an overall 99.86% confidence for the territory of the Netherlands (or with a 95% confidence for each land use category), if *X. fastidiosa* is present, the number of infected plants is below 0.4%; and
- with an overall 98.43% confidence for the territory of the Netherlands (or with a 53% confidence for each land use category), if *X. fastidiosa* is present, the number of infected plants is below 1%.

3.3.2.2. Option 2: NUTS 3 regions for agricultural areas, NUTS 2 regions for urban areas, forests and other areas

Table 13: Proportional allocation of samples in each epidemiological unit for Option 2 (NUTS 3 regions for agricultural areas, NUTS 2 regions for urban areas, forests and other areas) and for a fixed design prevalence scenario for the entire territory (1%)

Land use	Design prevalence (%)	NUTS level (Option 2)	Number of epidemiological units	Samples 99.999% confidence level	Allocation of samples	Samples 95% confidence level	Allocation of samples
Urban areas	1	2	12	544	544/12 = 46	137	137/12= 12
Forest	1	2	12	544	544/12 = 46	137	137/12 = 12
Other areas (semi-wild)	1	2	12	544	544/12 = 46	137	137/12 = 12
Agricultural	1	3	40	544	544/40 = 14	137	137/40 = 4
Total				**2,176**	**2,216**	**548**	**592**

Conclusion derived from Proportional allocation of samples in each epidemiological unit for Option 2 (NUTS 3 regions for agricultural areas, NUTS 2 regions for urban areas, forests and other areas) and for a fixed design prevalence scenario for the entire territory (1%)

Assuming that:

(i) in the 12 NUTS 2 regions of the Netherlands, the epidemiology of *X. fastidiosa* is similar in all the urban areas, in all the forests and in all the other areas; and

(ii) in the 40 NUTS 3 regions of the Netherlands, the epidemiology of *X. fastidiosa* is similar in all the agricultural areas,

after implementing this survey, should all the samples test negative, it could be concluded that:

- with an overall 99.999% confidence for the territory of the Netherlands (or with a 95% confidence for each land use category), if *X. fastidiosa* is present, the number of infected plants is below 1%; and
- with an overall 95% confidence for the territory of the Netherlands (or with a 53% confidence for each land use category), if *X. fastidiosa* is present, the number of infected plants is below 1%.

Table 14: Proportional allocation of samples in each epidemiological unit for Option 2 (NUTS 3 regions for agricultural areas, NUTS 2 regions for urban areas, forests and other areas) and for a variable design prevalence scenario depending on the land use (1% for urban, forest and other areas and 0.4% for agricultural areas)

Land use	Design prevalence (%)	NUTS level (Option 2)	Number of epidemiological units	Samples 99.999% confidence level	Allocation of samples	Samples 95% confidence level	Allocation of samples
Urban areas	1	2	12	544	544/12 = 46	137	137/12= 12
Forest	1	2	12	544	544/12 = 46	137	137/12 = 12
Other areas (semi-wild)	1	2	12	544	544/12 = 46	137	137/12 = 12
Agricultural	0.4	3	40	1361	1361/40 = 35	343	343/40 = 9
Total				**2,993**	**3,056**	**754**	**792**

Conclusion derived from Table 14:

(i) in the 12 NUTS 2 regions of the Netherlands, the epidemiology of *X. fastidiosa* is similar in all the urban areas, in all the forests and in all the other areas; and

(ii) in the 40 NUTS 3 regions of the Netherlands, the epidemiology of *X. fastidiosa* is similar in all the agricultural areas,

after implementing this survey, should all the samples test negative, it could be concluded that:

- with an overall 99.86% confidence for the territory of the Netherlands (or with a 95% confidence for each land use category), if *X. fastidiosa* is present, the number of infected plants is below 0.4; and
- with an overall 98.43% confidence for the territory of the Netherlands (or with a 53% confidence for each land use category), if *X. fastidiosa* is present, the number of infected plants is below 1%.

3.3.2.3. Conclusion

The concluding statements derived from the survey designs of Tables 11 and 13 are comparable for both options (similarly for Tables 12 and 14) and are based on the same sample size. However, the underpinning assumptions are substantially different. This shows that the conclusion of the survey always needs to be associated with the assumptions made on the homogeneity.

3.3.3. Proportional allocation of the inspection units to the host population in the epidemiological units

If the host plant population within each epidemiological unit is well known, the sample can be allocated accordingly. This approach can be applied for all land use categories. As an example, Table 15 shows the distribution of the samples in the urban areas of the four NUTS 1 regions in the Netherlands.

Table 15: Example of a proportional allocation of the samples to the host plant population in the epidemiological units

Land use	Overall sample size	NUTS1 code	NUTS 1 region	Proportion of host plant population[(a)] per NUTS 1 region	Sample size
Urban area	544	NL1	NOORD-NEDERLAND	0.60	544×0.6=327
		NL2	OOST-NEDERLAND	0.20	544×0.2=109
		NL3	WEST-NEDERLAND	0.05	544×0.05=28
		NL4	ZUID-NEDERLAND	0.15	544×0.15=82
Total				**1**	**546**

(a) Figures are for illustrative purposes only.

3.4. Risk-based detection survey

3.4.1. Risk locations

For *X. fastidiosa*, the production, trade and movement of host plants for planting are considered risk activities. The relevant risk locations are nurseries including tree nurseries and garden centres where these activities take place.

In 2020, in the Netherlands about 2,580 tree nurseries and 500 garden centres are identified. Table 16 shows how these locations are distributed in the NUTS 2 regions of the Dutch territory. The figures used in this example are not official data and some are estimated values.

Table 16: Example of distribution of different types of risk location in the NUTS 2 regions of the Netherlands

Code	NUTS 1	NUTS 2	Total nurseries[(a)]	Tree nurseries[(a)]	Garden centres[(a)]
NL1	**NOORD-NEDERLAND**				
NL11		Groningen	40	25	17
NL12		Friesland	31	19	19
NL13		Drenthe	62	39	14
NL2	**OOST-NEDERLAND**				
NL21		Overijssel	140	87	33
NL22		Gelderland	872	544	60
NL23		Flevoland	117	73	12
NL3	**WEST-NEDERLAND**				
NL31		Utrecht	209	130	39
NL32		Noord-Holland	158	99	83
NL33		Zuid-Holland	676	422	106
NL34		Zeeland	318	198	11
NL4	**ZUID-NEDERLAND**				
NL41		Noord-Brabant	1,026	640	74
NL42		Limburg	487	304	32
Total Netherlands			**4,136[(b)]**	**2,580**	**500**

(a) The distribution in the Dutch NUTS 2 regions of the risk locations are estimated values and are only an approximation.
(b) Total number of nurseries in the Netherlands in 2009.

3.4.2. Relative risk

Once the risk locations have been identified it is necessary to categorise them into levels and to determine their risks relative to the rest of the territory. To estimate these values, historical information on interceptions, trade volumes and the origins of the plant material can be considered.

There are three different types of risk location, based on their relative risks, which have been estimated using expert knowledge:

(i) **High-risk locations, with a relative risk of 2**: nurseries or garden centres that import plant material from countries where *X. fastidiosa* is known to occur. It is assumed that this relative risk applies to one third of the sites.

(ii) **Medium risk locations, with a relative risk of 1.5**: nurseries or garden centres that are located in an area where host plants are growing but that do not import plant material from countries where *X. fastidiosa* is known to occur. It is assumed that this relative risk applies to two thirds of the sites.

(iii) **Baseline, with a relative risk of 1**: all other areas where host plants are growing.

Risk area around each risk location

As indicated in Section 2.4.1, related to the spread capacity of the bacterium, the area to survey around each risk location should have a width of 150 m. Assuming that an average risk location covers 1 ha, the risk area to survey around the risk location is 16 ha (Figure 5).

Proportion of the host population for each risk location

A practical example has been developed for the survey design for the Netherlands, where 4,636 (4,136 + 500) risk locations were identified (see Table 16). From these, about one third (1,545) of the locations classified as high-risk locations and two thirds (3,091) classified as medium-risk locations.

The World Bank collection of development indicators (World Bank, online), compiled from officially recognised sources, provides the information used to calculate the total agricultural area of the Netherlands used in this example. The surface of the relevant agricultural area is calculated by deducting the land under cereal production (non-*X. fastidiosa* hosts) from the arable land in the Netherlands, which comes to 870,000 hectares.

The number of host plants that are in the epidemiological unit is estimated to be on average **100 host plants per hectare** of agricultural land. These are also the numbers used for the survey design in the example given below.

3.4.3. Integrating the risk factor into the survey design with RiBESS+

This section presents an example to illustrate the design of a risk-based detection survey for *X. fastidiosa* in the agricultural areas of the Netherlands. The survey parameters used in the example are summarised in Table 17.

Table 17: Example of relative risk and proportion of host plant population

Land use	Design prevalence (%)	Confidence level (%)	Method sensitivity	Risk factor				
					Number of risk locations	Agricultural land (hectare)[a]	Proportion of host plant population[b]	Relative risk[c]
Agriculture area in the Netherlands	0.4	**53**	0.55	High risk	1,545	24,720	2.8%	2
				Medium risk	3,091	49,456	5.7%	1.5
				Baseline	-	795,824	91.5%	1
				Total	4,636	870,000	100%	-

(a) 16 ha per risk location.
(b) 100 host plants per hectare, on average.
(c) Relative risk estimated with expert knowledge.

Figures 9–11 show how to input the various parameters and information. The first step of this survey design is to include the input parameters in the RiBESS+ tool (Figure 9). The second step is to introduce the risk factors (Figure 10). As summarised in the table above, one risk factor with three levels of risk is used. The third step (Figure 11) shows how other sampling schemes could be employed, such as convenience sampling (in which samples are distributed according to other criteria and the number of samples is allocated within each risk level). In this example, the surveyor decided to sample five times more in the high-risk locations than in the baseline areas and three times more in the medium locations than in the baseline areas.

The resulting sample size is calculated with the same confidence and design prevalence for each risk level as shown in Table 18.

There is a consequent reduction of sample size when comparing the survey design without using risk factors (343) with the survey design using risk factors (263).

When applying the convenience sampling approach, the sample size is further reduced (from 263 to 216).

Table 18: Sample size calculated using RiBESS+ for a detection survey

Land use	Design prevalence (%)	Confidence level (%)	Samples without risk factor	Samples with risk factor			
				Risk level	No convenience sampling	Convenience sampling	
Agriculture area	0.4	53	343		263		216
				High risk	**61**	**5**	**120**
				Medium risk	**81**	**3**	**72**
				Baseline	**121**	**1**	**24**

Step 1: calculate the sample size for the detection survey (Figure 9).

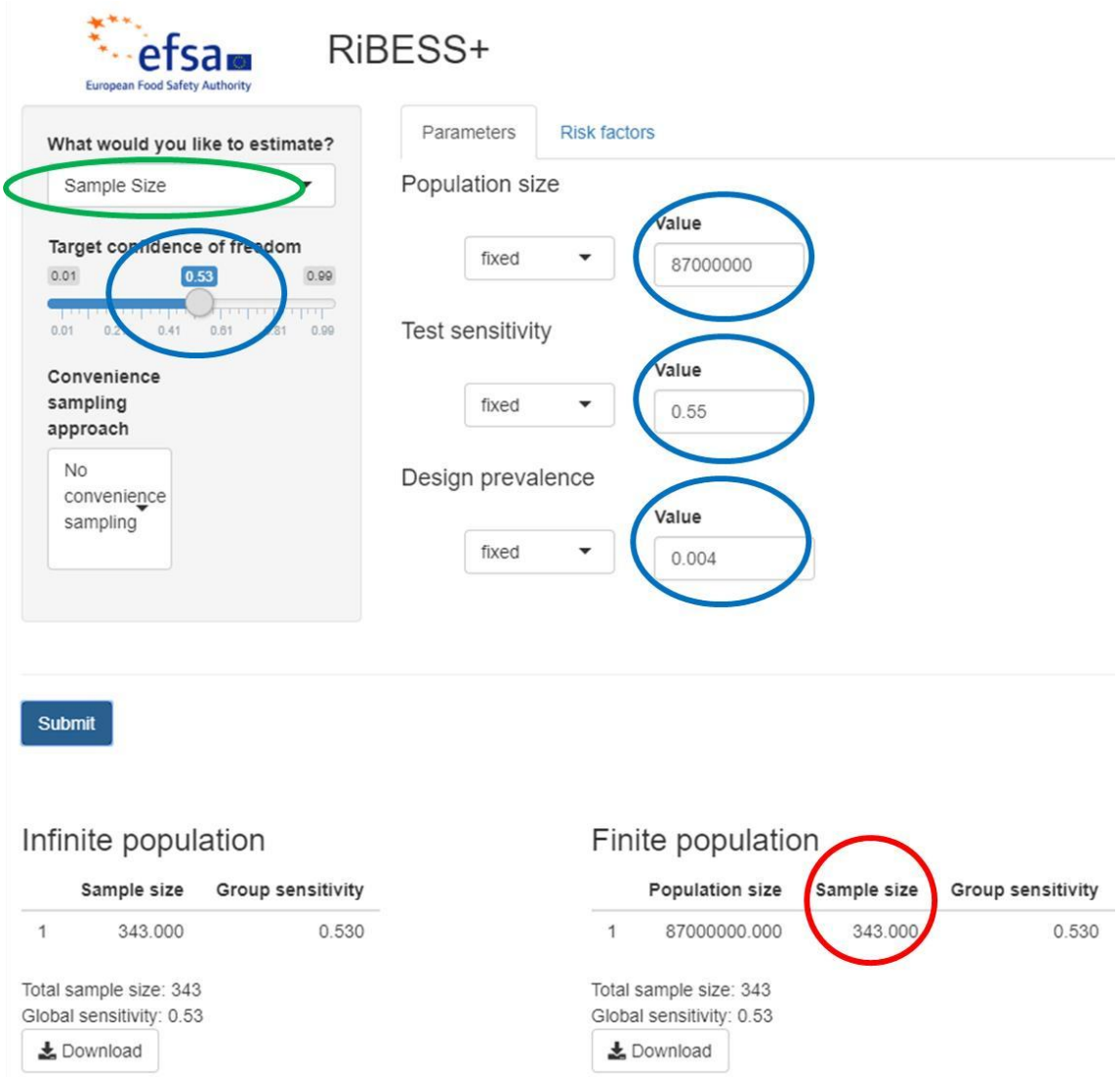


Figure 9: Screenshot of RiBESS+ illustrating the estimation of sample size for a survey (the green circle is the chosen functionality, the blue circles are the input values of the survey parameters, the red circle is the estimated output)

Step 2: integrate the risk factor with three levels into the survey design (Figure 10).

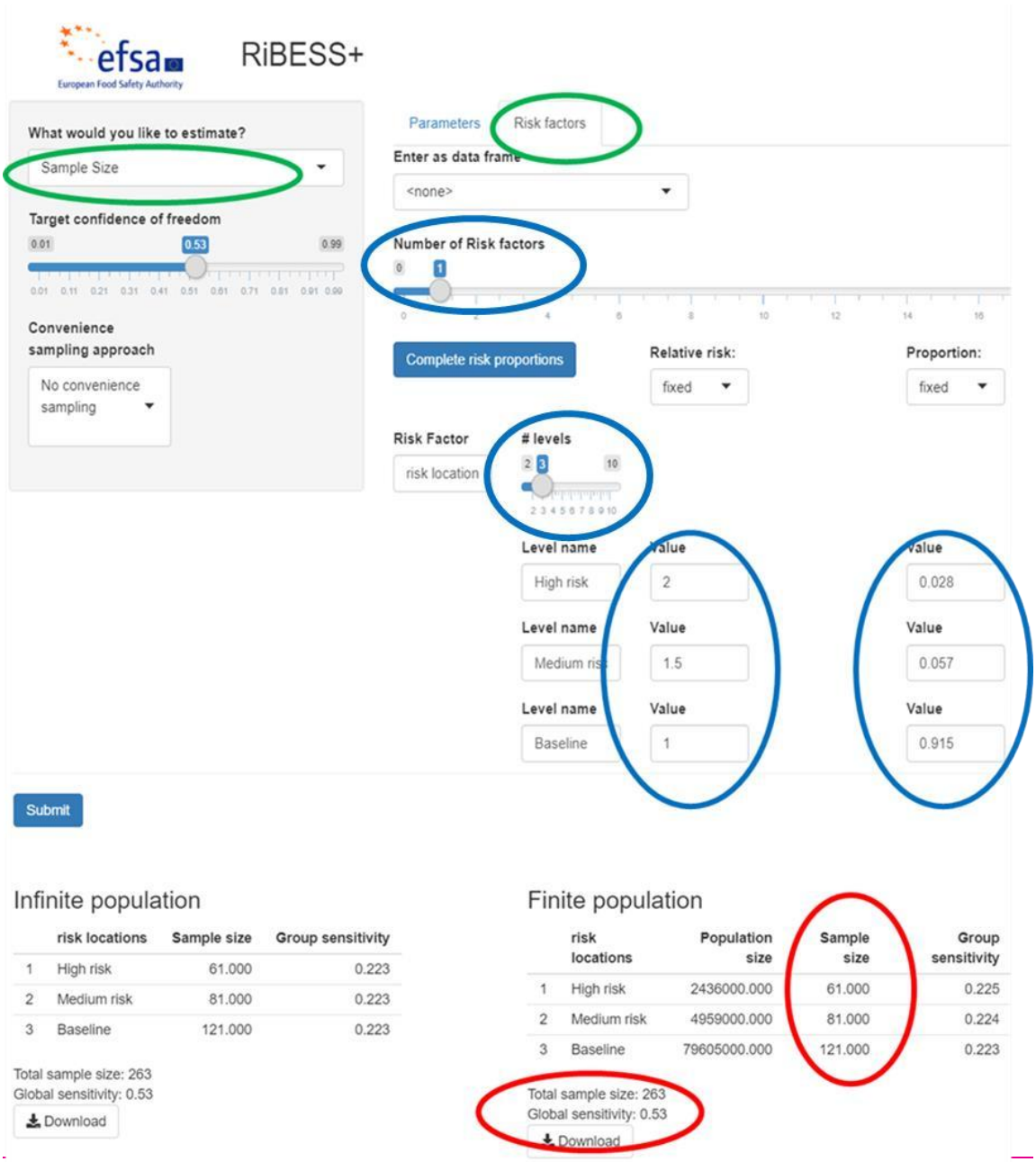


Figure 10: Screenshot of the integration of one risk factor with three levels into the survey design in RiBESS+ (the green circles are the chosen functionalities, the blue circles are the input values of the survey parameters, the red circles are the estimated output values)

Step 3: choose the sampling approach. In the example, a convenience sampling has been chosen (Figure 11).

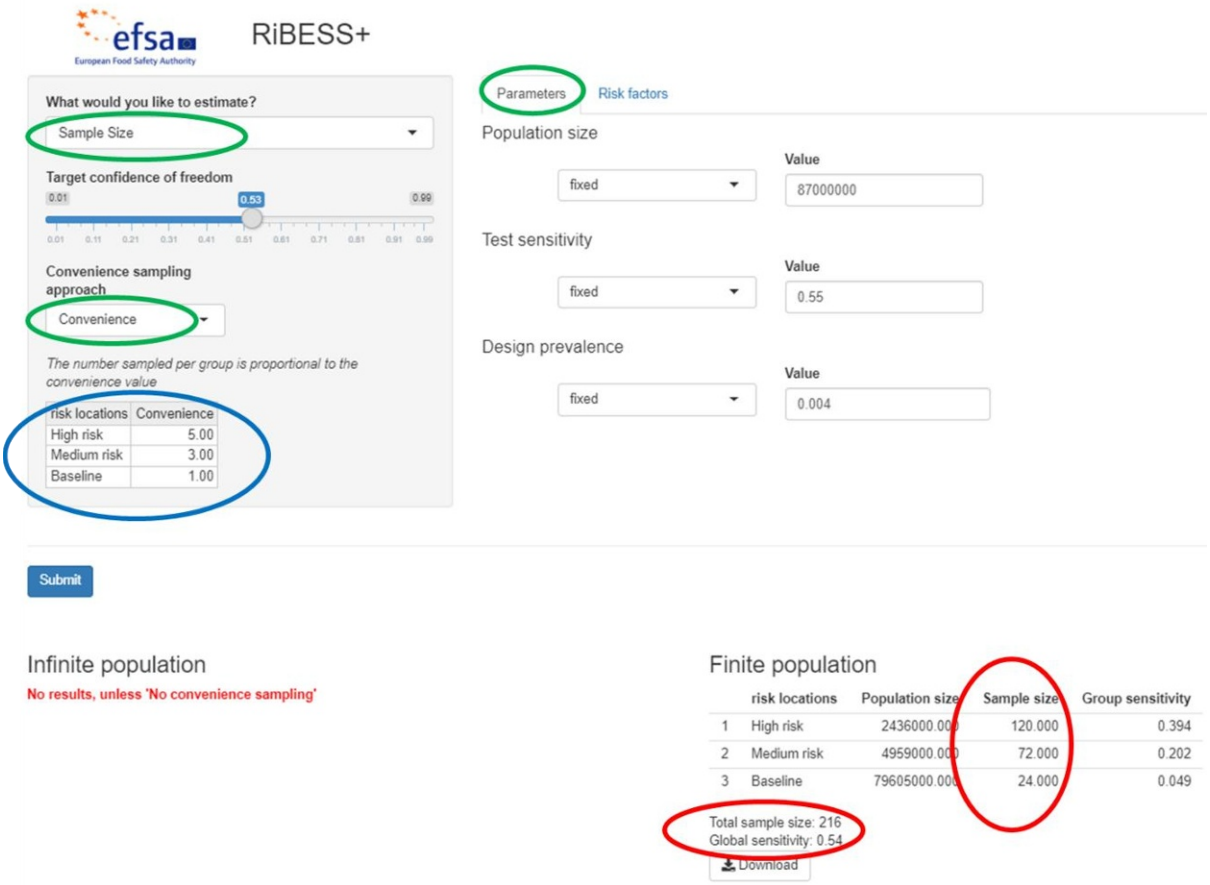


Figure 11: Screenshot of a convenience sampling calculation using RiBESS+ at three levels, i.e. five times more samples to be taken around the high-risk locations than in the baseline area and three times more samples to be taken around the medium-risk locations than in the baseline area (the green circles are the chosen functionalities, the blue circle is the input values of the survey parameters, the red circles are the estimated output values)

3.4.4. Proportional allocation of the samples to the number of risk locations

The allocation of the samples estimated for the three risk levels using a convenience sampling approach for the Netherlands can be done proportionally to the risk locations that are in each NUTS 2 region. The 24 remaining samples that need to be taken in the baseline area can be allocated equally in the 12 NUTS2 regions. The results of these calculations are shown in the table below (Table 19).

Table 19: Example of a proportional allocation of the samples to the number of risk locations in the NUTS 2 regions of the Netherlands

Code	NUTS 1	NUTS 2	Total nurseries[(a)]	Garden centres[(a)]	High-risk locations			Medium risk locations			Baseline
					N	Proportion	Samples	N	Proportion	Samples	Samples
NL1	**NOORD-NEDERLAND**										
NL11		Groningen	40	17	19	0.0041	1	38	0.0082	1	2
NL12		Friesland	31	19	17	0.0036	1	33	0.0072	1	2
NL13		Drenthe	62	14	25	0.0055	2	51	0.0109	1	2
NL2	**OOST-NEDERLAND**										
NL21		Overijssel	140	33	58	0.0124	4	115	0.0249	3	2
NL22		Gelderland	872	60	311	0.0670	24	621	0.1340	14	2
NL23		Flevoland	117	12	43	0.0093	3	86	0.0186	2	2
NL3	**WEST-NEDERLAND**										
NL31		Utrecht	209	39	83	0.0178	6	165	0.0357	4	2
NL32		Noord-Holland	158	83	80	0.0173	6	161	0.0347	4	2
NL33		Zuid-Holland	676	106	261	0.0562	20	521	0.1125	12	2
NL34		Zeeland	318	11	110	0.0237	9	219	0.0473	5	2
NL4	**ZUID-NEDERLAND**										
NL41		Noord-Brabant	1,026	74	367	0.0791	28	733	0.1582	17	2
NL42		Limburg	487	32	173	0.0373	13	346	0.0746	8	2
Total Netherlands			**4,136**	**500**	**1,545**	**0.3333**	**120**	**3,091**	**0.6667**	**72**	**24**

3.5. Estimation of the confidence of an implemented survey

Following the implementation of the survey, it is possible to calculate the achieved or realised confidence level of the survey under the same assumptions using the same methods for detection and identification.

Figure 12 below illustrates an example using the RiBESS+ tool to calculate the achieved confidence of the survey **without including risk factors**. After selecting the option for estimating the 'global sensitivity' or confidence level, the survey parameters should be provided in the tool. For example, if 130 samples were collected in agricultural areas and tested negative for *X. fastidiosa*, considering a target population of 87,000,000 host plants in the survey area, a method sensitivity of 0.55 and a design prevalence of 0.4%, then the calculated confidence level of that survey is 25%.

If all 130 samples tested negative, as a conclusion of this survey, under the assumption of homogeneity of the surveyed areas, it could be stated with 25% confidence that, if *X. fastidiosa* is present in the agricultural areas, it is below 0.4% prevalence (infected host plants). If in this case the survey was designed to conclude with 53% confidence for the agricultural areas, then the objective has not been met.

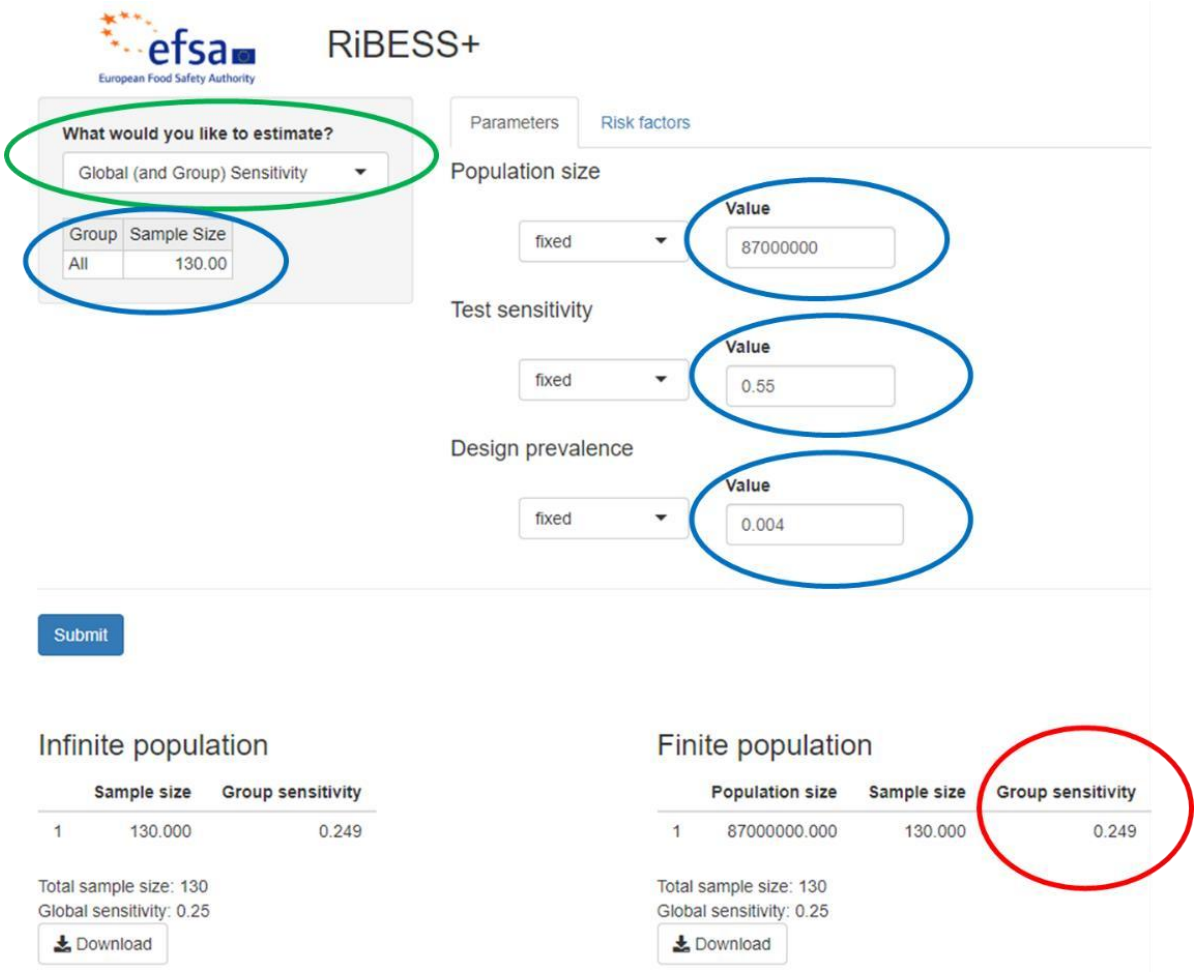


Figure 12: Screenshot of RiBESS+ illustrating the estimation of the achieved confidence level of a survey without risk factors (the green circle is the chosen functionality, the blue circles are the input values of the survey parameters, the red circle is the estimated output value)

Figure 13 illustrates an example using the RiBESS+ tool to calculate the achieved confidence of the survey **including risk factors**. After selecting the option for estimating the 'global sensitivity' or confidence level, the survey parameters should be provided in the tool. Then 'risk factor' should be selected and filled in as shown in the previous section. For example, if in agricultural areas 70 samples

were collected in high-risk areas, 50 samples in medium-risk areas and 10 in the baseline, and all tested negative for *X. fastidiosa*, then considering a target population of 87,000,000 host plants in the survey area, a method sensitivity of 0.55 and a design prevalence of 0.4%, the calculated confidence level of that survey is 37%.

If in this case the survey was designed to conclude with 53% confidence for the agricultural area then the objective has not been met.

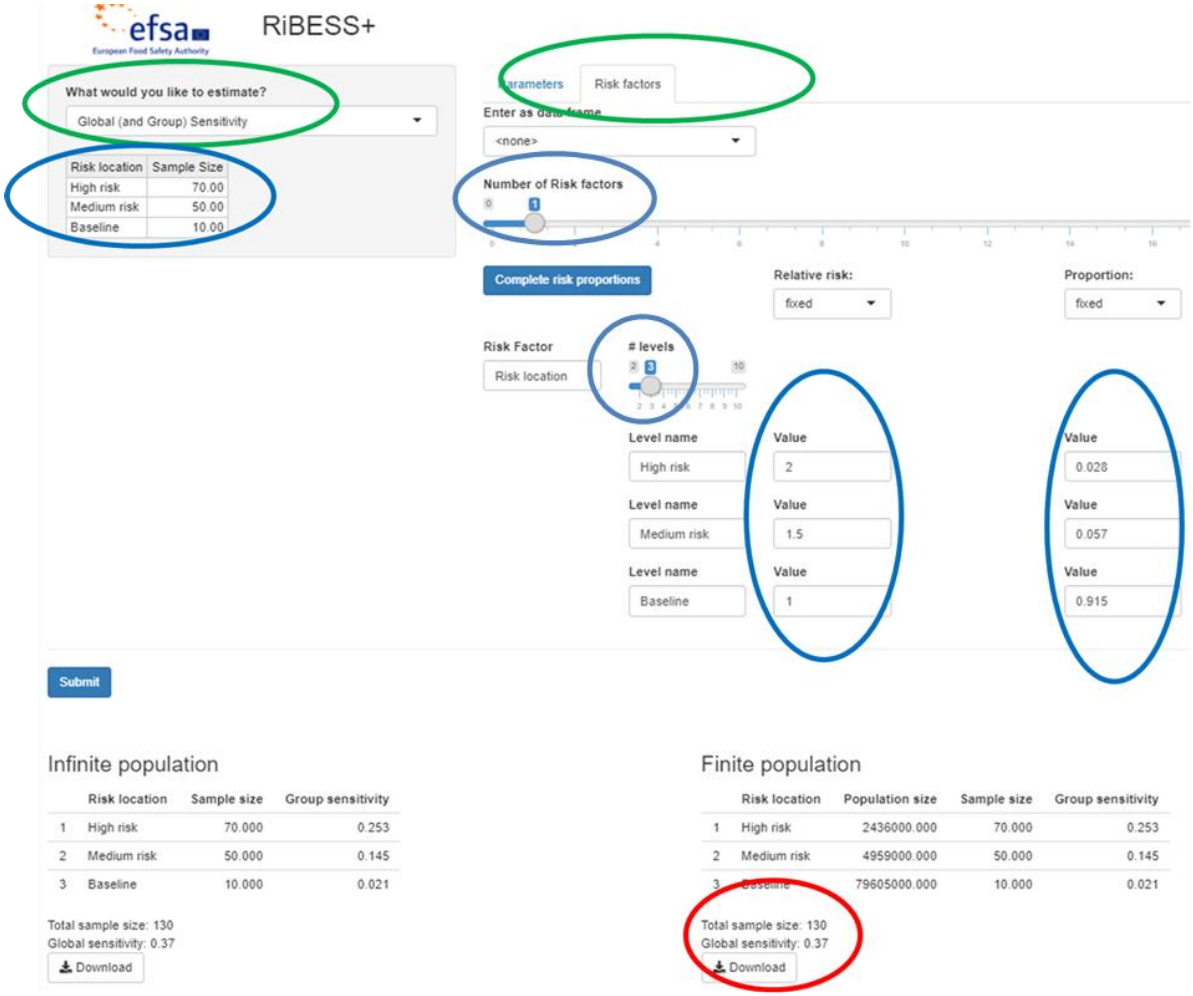


Figure 13: Screenshot of RiBESS+ illustrating the estimation of the achieved confidence level of a survey with risk factors (the green circles are the chosen functionalities, the blue circles are the input values of the survey parameters, the red circle is the estimated output value)

It is to be noted that both simulations result in a different confidence level: 25% without considering risk factors and 37% with them. The more information is assembled in the survey design, the better the survey can be targeted and the higher the confidence will be.

4. Delimiting survey

A delimiting survey is conducted to establish the boundaries of an area considered to be infested with or free from a pest (International Standards for Phytosanitary Measures, ISPM 5 (FAO, 2020)).

The delimiting survey is performed after a positive finding of *X. fastidiosa* to delimit an area or infested zone where the pest is contained.

Within the delimited area or infested zone, an eradication or containment programme (ISPM 9, FAO, 2016) will be implemented, but this is not addressed in these guidelines.

Around the infested zone, a buffer zone is established and intensively surveyed to protect the territory from further spread of the pest and to ensure pest freedom over time. The infested zones with the buffer zone are defined in the EU legislation as the 'demarcated area'[7].

However, immediately after the first finding is confirmed, a provisional demarcation is performed where official measures are applied to avoid further spread of the disease. The results of the delimiting survey will confirm the definitive boundaries of the demarcated area.

In this section, hypothetical scenarios are given for illustration purposes.

4.1. Delimiting survey strategy

The delimiting survey strategy is based on a sequence of detection surveys performed in peripheral bands going inwards to the centre of the potentially infested zone. The width of the survey bands is defined according to the yearly local spread capacity of the disease.

Within each peripheral band, the freedom from *X. fastidiosa* can be defined within a statistical framework, with 95% confidence that the prevalence of the bacterium in the surveyed target population, if present, is below the design prevalence. A stepwise procedure is described below.

4.1.1. Step 1. Find the source of infection after a positive finding of *Xylella fastidiosa*

Once *X. fastidiosa* has been detected in an area, a tracing back and forward procedure should be conducted by the NPPO following their standard operating procedures as required by ISPM 9 (FAO, 2016), to identify other potential locations that could have been exposed to similar infection, and locations that could have been exposed following the further spread of the bacterium. Intensive monitoring in all the identified areas should be carried out.

Following the finding of an infected plant, the most plausible infection source should be identified by monitoring the risk locations within a natural spread distance. The locations found to be infected are considered to be the source of the infection. Different cases can be distinguished as shown in Figure 14.

[7] Regulation (EU) 2016/2031 of the European Parliament of the Council of 26 October 2016 on protective measures against pests of plants, amending Regulations (EU) 228/2013, (EU) 652/2014 and (EU) 1143/2014 of the European Parliament and of the Council and repealing Council Directives 69/464/EEC, 74/647/EEC, 93/85/EEC, 98/57/EC, 2000/29/EC, 2006/91/EC and 2007/33/EC. OJ L 317, 23.11.2016, pp. 4–104.

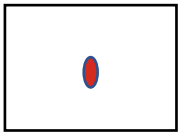
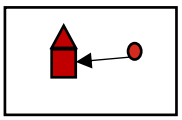
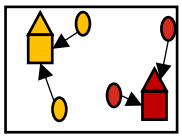
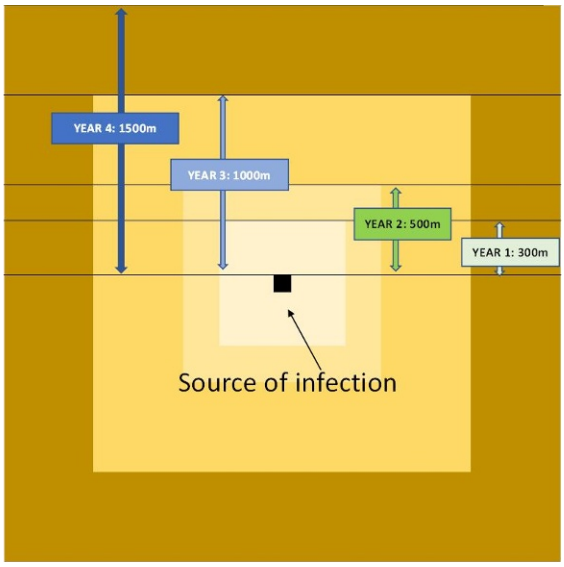
One infected host was found; no source locally identified. The infected host becomes the source of the infection.	
One infected host was found and a source identified.	
More than one infected host was found; no sources locally identified. All infected hosts are considered as sources.	
More than one source of infection was identified.	

Figure 14: Different scenarios for the identification of the source of an infected host plant

4.1.2. Step 2. Estimate the boundaries of the potentially infested zone

Around the source of the infection the potentially infested zone must be determined and its extension depends on the time and rate at which the pest has been spreading in the area. The yearly local disease dispersal distances specified in EFSA PLH Panel (2019) show that the spread accelerates with time. Therefore, the width of the area can be assumed to correspond to the distance the disease has spread since the last negative detection survey in the area. Figure 15 indicates the distance to consider around the source of the infection of *X. fastidiosa* for a delimiting survey, depending on the time elapsed since the last detection survey was conducted.



Years since last detection survey of the site	Estimated spread distance around the source of infection [a]
1	300 m [b]
2	500 m [c]
3 [d]	**1000 m** [c]
4	1500 m [c]

(a) The potential spread distance, from its introduction until the pest is n found, in the worse case corresponds to the years elapsed since the last detection survey was performed.
(b) Yearly median of short-distance dispersal 151 m (fitted to the spread rate in Apulia) (EFSA PLH Panel, 2019).
(c) Based on short-range spread model of the disease caused by *Xylella fastidiosa* (EFSA PLH Panel, 2019).
(d) This is the scenario chosen for the simulations.

Figure 15: Potentially infested zone width to consider around the source of infection for a delimiting survey of *Xylella fastidiosa* depending on the time elapsed since the last detection survey was conducted, taking into account the accelerating disease spread distances (EFSA PLH Panel, 2019)

In EFSA PLH Panel (2019), the maximum short-distance disease dispersal fitted to the Apulian epidemics was estimated to be below 400 m. Therefore, a band 400 m wide around the potentially infested zone is defined for performing the survey.

Figure 16a shows the potentially infested zone (PIZ 1) and the additional 400 m band (Band 1) under the assumption that the last detection survey was carried out 3 years ago. Figure 16b illustrates the situation where two potentially infested zones overlap for the delimiting survey. The procedure as developed in this section can also be applied in this case.

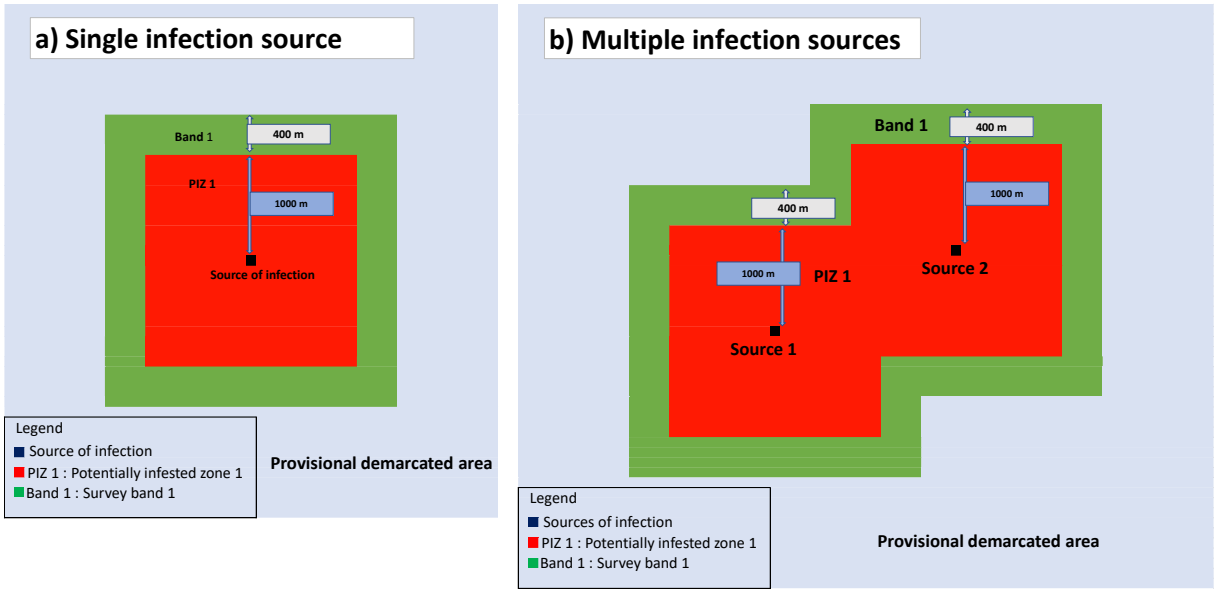


Figure 16: Graphical description of the potentially infested zone around the source of infection and the additional 400 m surrounding the potentially infested zone (i.e. Band 1) where the survey should be conducted first. (a) Single infection source was identified. (b) Multiple infection sources were identified.

4.1.3. Step 3. Delimit the boundaries of the infested zone

This step starts with surveying Band 1. After analysing the samples collected, two different situations could arise:

4.1.3.1. Step 3a. Narrowing down the potentially infested zone

If all the samples taken in Band 1 have tested negative to *X. fastidiosa* then Band 1 is cleared. An inner Band 2 is defined for a new round of the survey. The potentially infected zone 1 (PIZ 1) is narrowed down by 400 m from its periphery, defining Band 2 and PIZ 2 (Figure 17). After analysing the samples collected, again two different situations could arise:

1. If all the samples taken in Band 2 have tested negative to *X. fastidiosa* then Band 2 is cleared. An inner Band 3 is defined for a new round of the survey. PIZ 2 is narrowed down by 400 m from its periphery, defining Band 3 and PIZ 3. This process is repeated until the source of infection is reached.
2. If at least one infected plant was found in Band 2, the boundaries of the infested zone are confirmed, and the infested zone is delimited as PIZ 2 + Band 2. Around the infested zone a buffer zone of 10,000 m is defined. This is the median value of the long-distance dispersal fitted to the Apulian data (EFSA PLH Panel, 2019). This is shown in Figure 17. The buffer zone should be surveyed yearly to ensure it is kept free of the pest.

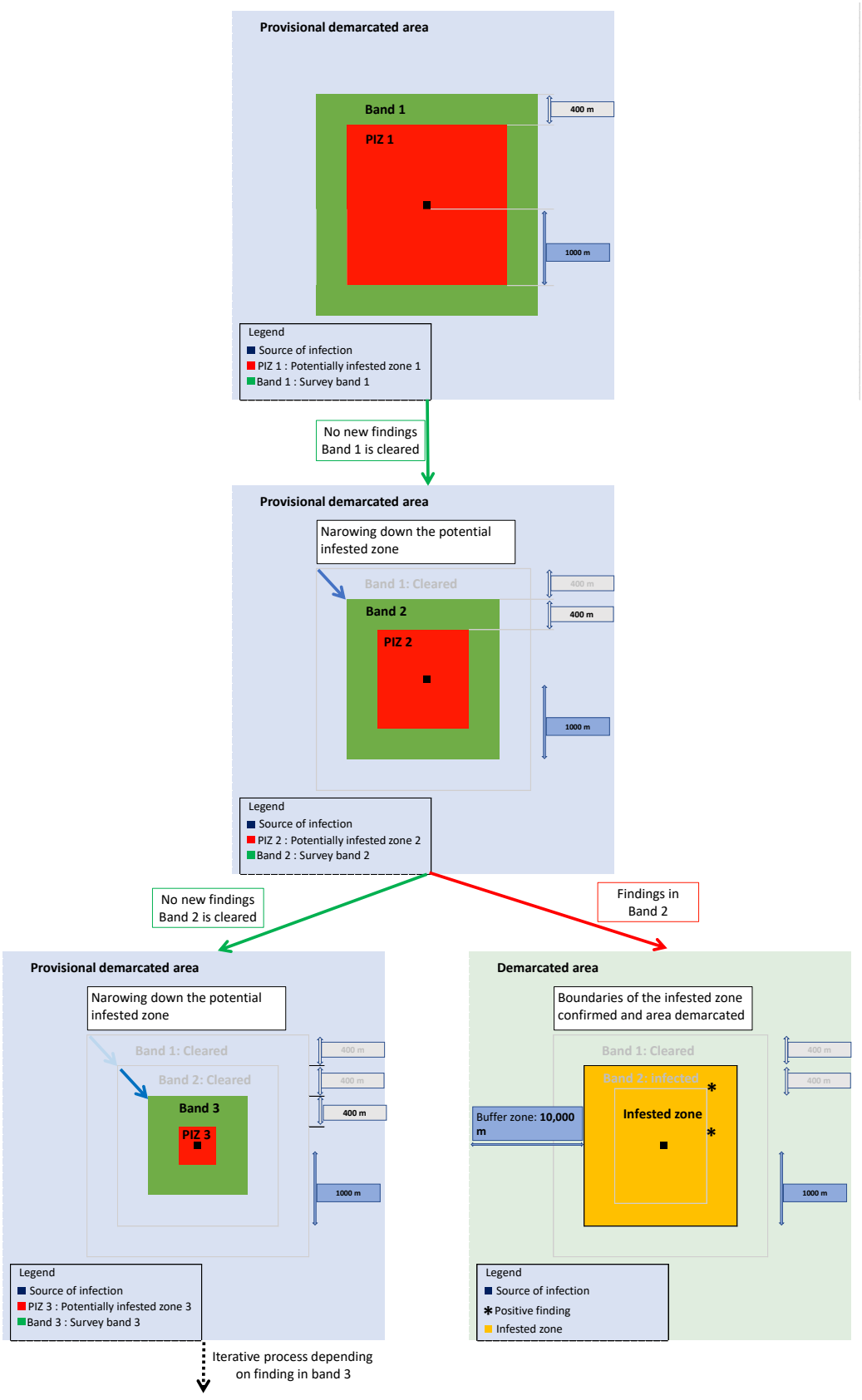


Figure 17: Graphical description of the Step 3a to confirm the boundaries of the infested zone. A buffer zone of 10 km is defined around the infested zone to define the demarcated area

4.1.3.2. Step 3b. Enlarging the potentially infested zone

If at least one sample tested positive to the bacterium in Band 1, PIZ 1 is enlarged to include Band 1 defining PIZ 2. An additional 400 m surrounding PIZ 2 should be surveyed, i.e. Band 2, using the same design prevalence and confidence level and method sensitivity as the survey for Band 1. This process is iterative until one band is found free of the pest and is cleared. This is illustrated in Figure 18. When a survey band of 400 m is found to be free of the pest, the delimitation process is finalised, and the demarcated area is defined as the confirmed infested zone surrounded by a buffer zone of 10 km.

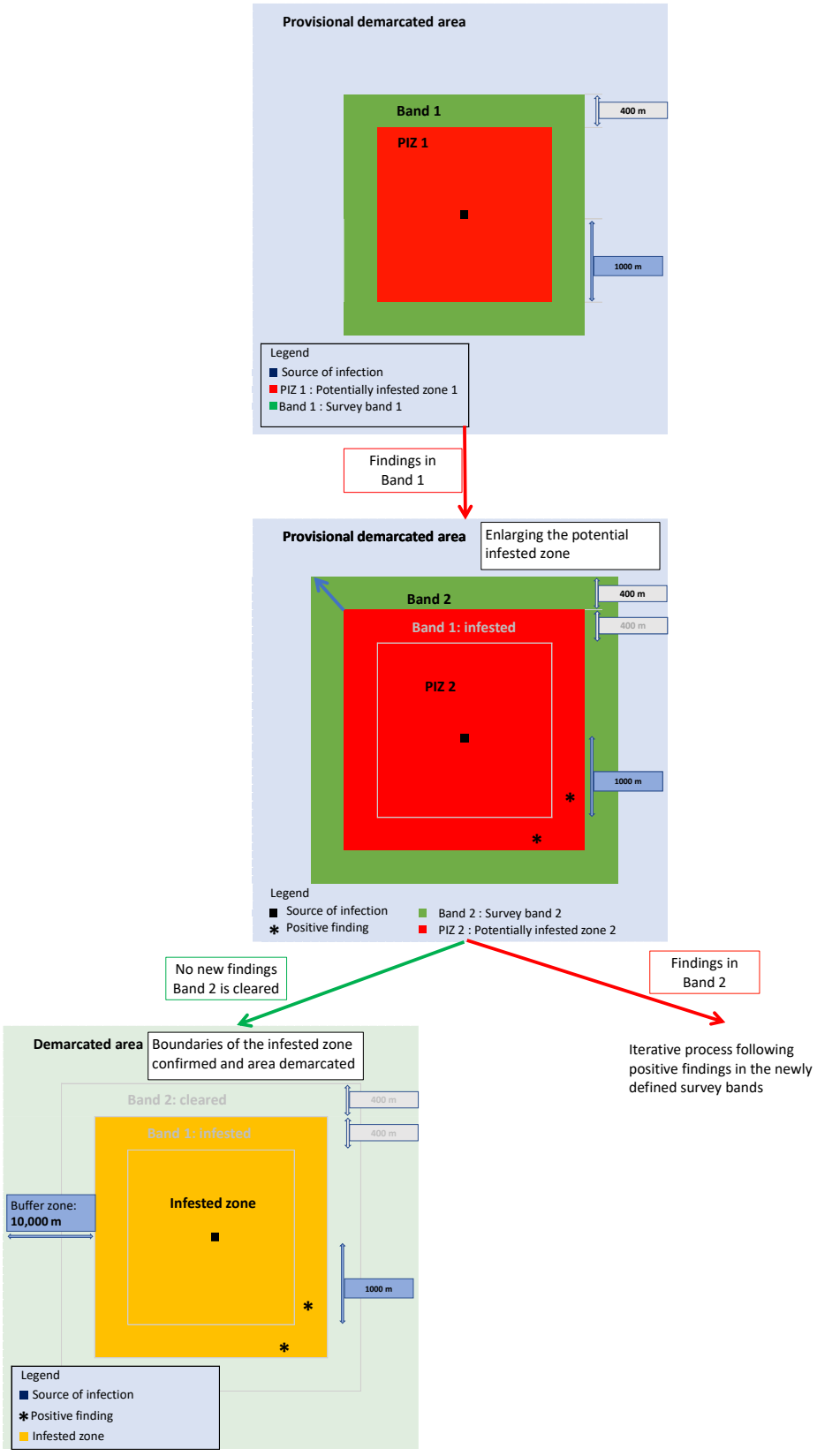


Figure 18: Graphical description of Step 3b. At least one infected plant tested positive in the first peripheral band of 400 m (Band 1) around the potentially infested zone 1; an additional peripheral band to survey (Band 2) is defined around the enlarged potentially infested zone 2 (Band 1 + potentially infested zone 1)

4.2. Survey parameters

4.2.1. Example definition

The survey parameters are estimated for the following example.

'An infected plant has been found in an agricultural area. The source of the infection has been identified as a glasshouse company extended over 1 ha. Three years ago, a detection survey was conducted in the same area, and no infected plants were found. It was estimated that each hectare has on average 150 host plants. A provisional demarcation of the area was done.'

The survey should start in Band 1 (Figure 19) and if at least one sample tests positive in Band 1 the calculation should be conducted similarly for the next band.

4.2.2. Target population and epidemiological unit

The entire survey band, defined in Step 2 of the delimiting survey strategy, is considered as a single epidemiological unit: the 400 ha to survey are assumed to be entirely within a homogeneous agricultural area.

The area of the first 400 m peripheral band (Band 1) should be estimated as shown in Figure 19.

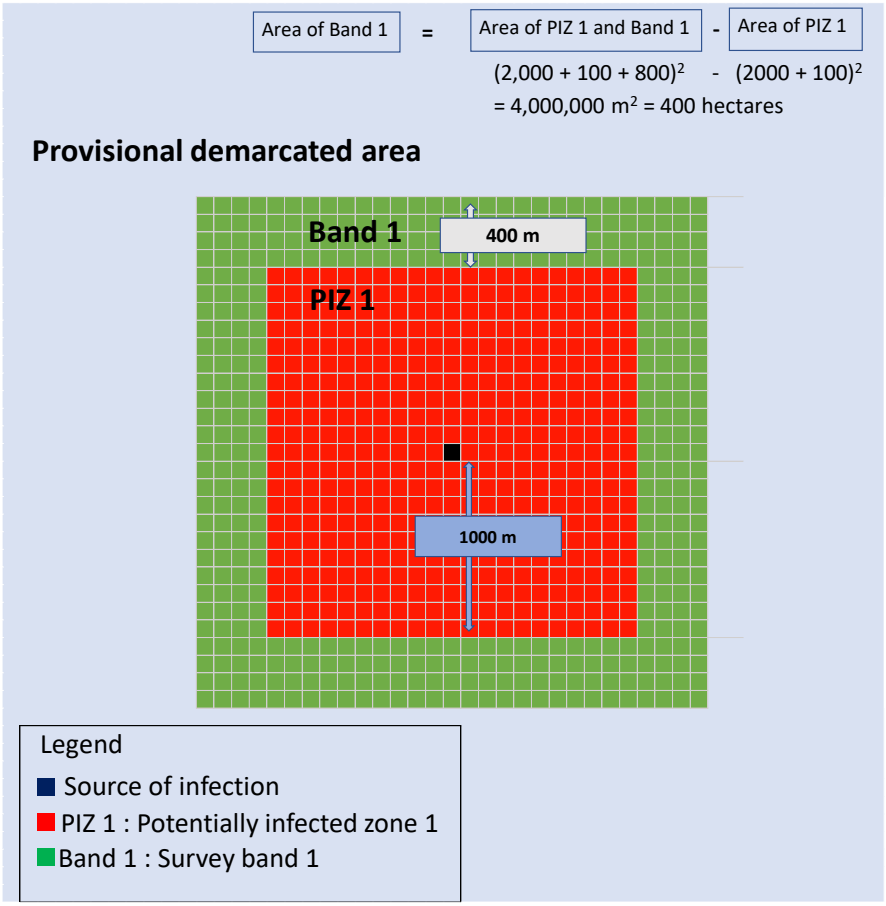


Figure 19: Calculation of the hectares of Band 1 for a delimiting survey to be conducted 3 years after the last detection survey in the area

Following the same reasoning, the number of hectares corresponding to the 400 m wide band for each year since the last detection survey can be calculated as shown in Table 20.

Table 20: Area (in hectares) to include in the first peripheral band of 400 m around the potentially infested zone for a delimiting survey of *Xylella fastidiosa*, depending on the time elapsed since the last detection survey was conducted and taking into account the accelerating disease dispersal distances (EFSA PLH Panel, 2019)

Years since last detection survey of the site	Potentially infested zone around the source of infection	Area (ha) of the first peripheral band of 400 m around the potentially infested zone
Year 1	300 m	176
Year 2	500 m	240
Year 3[(a)]	**1,000 m**	**400**
Year 4	1,500 m	560

(a) This is the scenario chosen for the simulations.

4.2.3. Confidence level and design prevalence

The aim of the survey designed in this section is to delimit the area where *X. fastidiosa* is contained following a positive finding with 95% confidence that if the bacterium is present in the area, the number of infected host plants is below the specified design prevalence. In Section 2.5, Table 5 shows an example of the design prevalence chosen for delimiting and buffer zone surveys compared with detection surveys for pest freedom substantiation and the rationale behind this choice is provided. As an example, the values used in this section are the following:

- design prevalence of 0.04% for a delimiting survey in agricultural areas (compared with 0.4% for an annual detection survey)
- design prevalence of 0.1% for a delimiting survey in urban areas, forests and other areas (compared with 1% for an annual detection survey).

4.2.4. Summary table for survey parameters

The survey parameters are summarised in Table 21.

Table 21: Survey parameters for an example of a *Xylella fastidiosa* delimiting survey in an agricultural area

Survey parameters	
Epidemiological unit: - Host density 150 plants/ha - 400 ha for Band 1	60,000 plants
Design prevalence for agricultural areas	0.04%
Confidence level	95%
Method sensitivity	0.55

4.3. Simulations

4.3.1. Sample size calculation

The sample size is 12,801, obtained using RiBESS+ as shown in Figure 20.

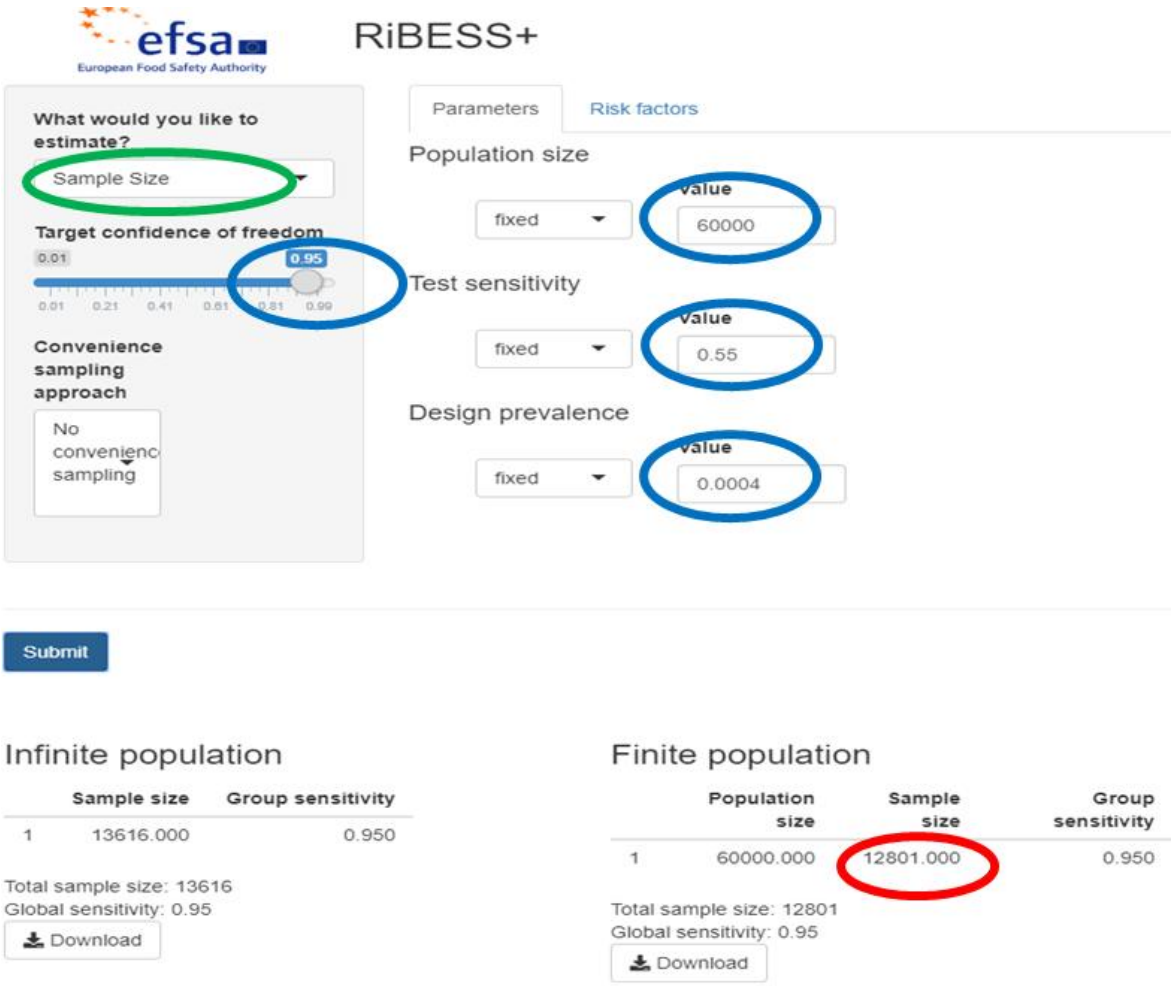


Figure 20: Screenshot of the sample size calculation using RiBESS+ for a detection survey of *Xylella fastidiosa* (the green circle is the chosen functionality, the blue circles are the input values of the survey parameters, the red circle is the estimated output value)

4.3.2. Sample allocation

If the samples are allocated proportionally across the entire area, this would correspond to 32 samples that need to be tested for each hectare in Band 1.

Within each hectare, the choice of the samples can be prioritised according to the probability of the host plant genus becoming infected, as determined in Appendix B.

4.4. Multiple infections

When multiple infections are found in an area, and these infections are not independent of each other, i.e. they result from the natural spread of the disease from the same initial introduction of the bacterium, it is recommended that they are grouped into a single infested zone. The infested zone is defined in Article 18(2) of the Plant Health Regulations 2016/2031 as the area comprising:

- all infected plants;

- all plants showing signs or symptoms;
- all other plants:
    - liable to have been or to become contaminated or infected due to:
        - their susceptibility

      and
        - their close proximity to infected plants

      or
        - their common source of production with infected plants

      or
        - their descent from infected plants;
- land, soil, watercourses or other elements infested, or liable to become infested.

This definition of an infested zone implies that if two infections are not independent, they should be combined into a single wider infested zone. One procedure that could be used to group them would be to identify the outer points at which infections were localised and join them with a straight line, creating a polygon that covers all local infections identified. A similar process to the one used to define the buffer and infested zones using the survey band procedure explained above could be followed.

Overall, when combining the infested zones, the survey sample for the delimiting survey will be reduced and the allocation of the samples in the area can be proportional to the host plant distribution.

5. Buffer zone surveys

In ISPM 5 (p. 12, FAO, 2020) a buffer zone is defined as 'An area surrounding or adjacent to an area officially delimited for phytosanitary purposes in order to minimize the probability of spread of the target pest into or out of the delimited area, and subject to phytosanitary or other control measures, if appropriate.'

Once the boundaries of the infested zone are established, a buffer zone of 10 km is defined around it. This is the median value of the long-distance dispersal fitted to the Apulian data (EFSA PLH Panel, 2019). Intensive surveillance is needed in the buffer zone to ensure the pest remains contained within the infested zone where an eradication programme is implemented. If such a survey finds infected plants in the buffer zone, delimiting surveys should be conducted to establish the new boundaries of the infested zone.

5.1. Survey parameters

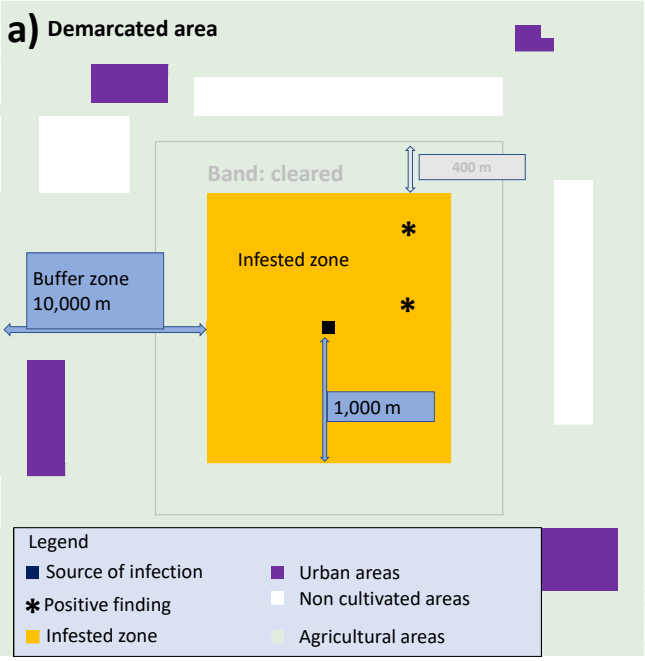
5.1.1. Target population

Considering the extent of the buffer zone of 10 km around the infested zone, it is important to consider the different land uses in the area. The survey should be conducted in the different environments and combined by means of the method presented in Section 2.5.

In our example, we consider only two different land use categories where host plants grow: agricultural and urban areas. All the urban areas are assumed to be homogeneous in terms of the epidemiology of *X. fastidiosa*. All the agricultural areas of the buffer zone are also considered homogeneous.

When selecting the host plants to be inspected, a prioritisation can be applied similar to that used for the delimiting survey (i.e. according to the ranking of the host plants by their probability of infection; see Appendix B) and this should be included in the inspection procedure.

The simulations are based on the following scenario (Figure 21a). The extent of the buffer zone is calculated as shown in Figure 21b.



b)

Area of Buffer zone = Area of Demarcated area - Area of infested zone

$(10,000+10,000+1,000+1,000+1)^2 - (1,000+1,000+1)^2$

= 48,004 hectares

Figure 21: (a) Schematic representation of the buffer zone; (b) Calculation of the area covered by the buffer zone

The host population in the buffer zone is structured as shown in Table 22.

Table 22: Host populations and area, per land use category in the buffer zone

	Hectares	**Host plants/ha**	**Host plant population**
Agricultural area	25,000	300	7,500,000
Urban area	12,000	100	1,200,000
Inaccessible areas or areas without host plant (e.g. rocky areas, roads)	11,000	N/A	N/A
Buffer zone	**48,000**		**9,300,000**

5.1.2. Confidence and design prevalence

The objective of the survey in the buffer zone is to ensure the area remains free from the pest and to detect *X. fastidiosa* at the very early stages of its introduction. Therefore, the aim is to detect the bacterium in the area with an overall confidence level of 95% that it is below a design prevalence accepted by the risk managers. In this example, as for the delimiting survey, the simulations will be

performed with the design prevalence of 0.04% for agricultural areas and 0.1% for urban areas. To achieve a survey with 95% confidence for the entire buffer zone, applying the method explained in Section 2.5, 78% confidence should be achieved for the agricultural areas and for the urban areas.

The parameters for the survey of the buffer zone are summarised in Table 23.

Table 23: Parameters used to design the survey of the buffer zone

Buffer zone survey	Agricultural areas	Urban areas
Confidence level	78%	78%
Design prevalence	0.04%	0.1%
Method sensitivity	0.55	0.55
Target population	7,500,000	1,200,000

5.1.3. Risk factor

For the delimiting survey, the outer band, 400 m wide, surrounding the infested zone, which was surveyed and found free from *X. fastidiosa*, can be considered to have a higher risk than the rest of the buffer zone. In our example, the first 400 m (400 ha) (high-risk area) are only in an agricultural area, and the host plants are considered to have double the risk of being infected as the remaining 9,600 m of the buffer zone (baseline).

If less than 1 year has elapsed between the survey conducted in the outer band and the demarcation of the buffer zone, then the samples analysed can be considered as evidence of pest freedom in the buffer zone and thus be deducted from the overall estimated survey sample.

In this example a convenience sampling is applied with twice as many samples taken in the high-risk area as in the baseline area.

The risk factor parameters for the simulation is summarised in Table 24 below.

Table 24: Risk factor parameters for survey of the buffer zone

Risk factor	Host plant population	Proportion of population	Relative risk	Convenience sampling
High risk	120,000	0.016	2	2
Baseline	7,380,000	0.984	1	1
Host plant population	7,500,000	1		

5.2. Simulations in the buffer zone

Figure 22 schematises the different areas within the demarcated area and in particular shows the risk areas of the buffer zone.

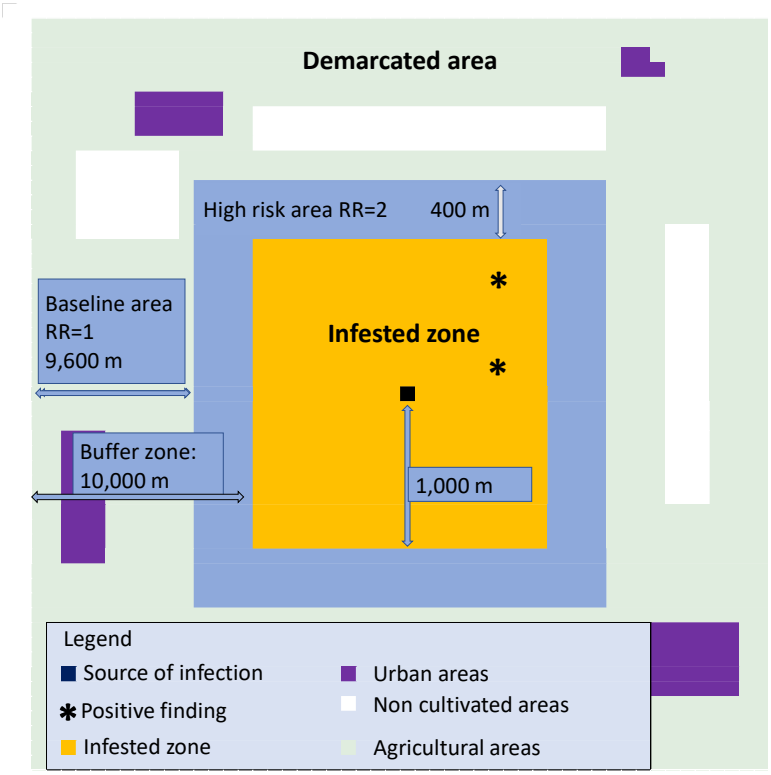
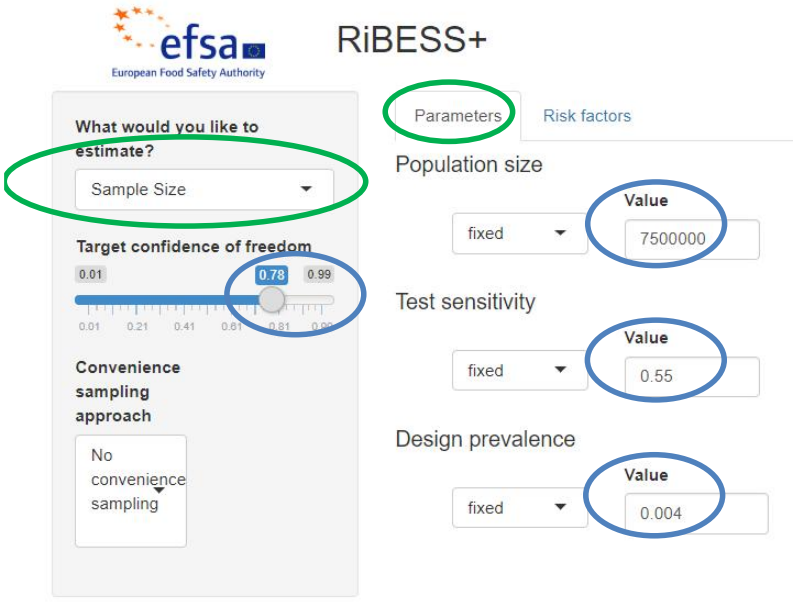


Figure 22: Schematic representation of the risk areas of the buffer zone

The simulations are presented separately for the agricultural area (Figure 23) and for the urban area (Figure 24).

5.2.1. Simulation for the agricultural areas of the buffer zone

First step:



Second step:

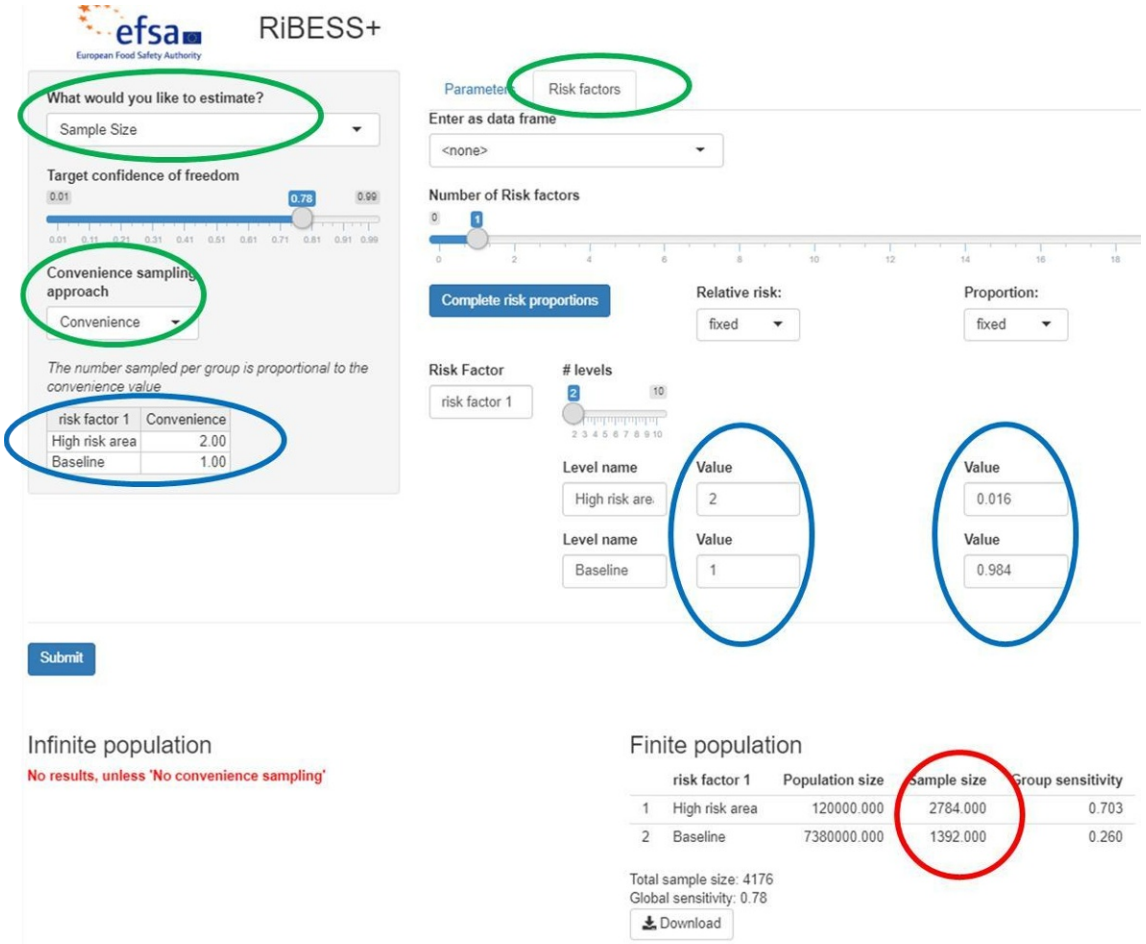


Figure 23: Screenshot of RiBESS+ for the calculation of the risk-based sample size for the agricultural areas of the buffer zone (the green circles are the chosen functionalities, the blue circles are the input values of the survey parameters, the red circle is the estimated output value)

5.2.2. Simulation for the urban areas of the buffer zone

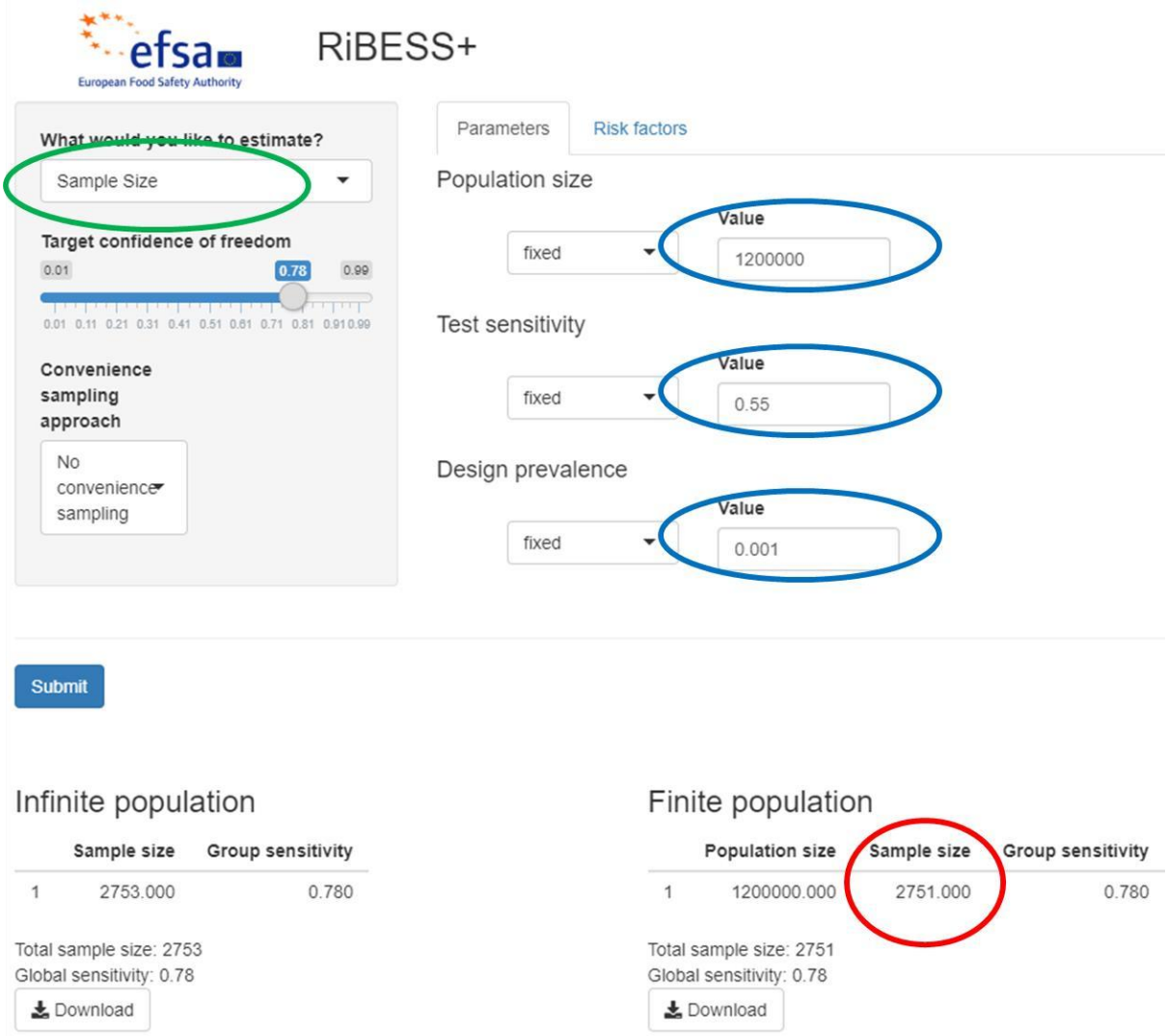


Figure 24: Screenshot of RiBESS+ for the calculation of the sample size for the urban areas of the buffer zone (the green circle is the chosen functionality, the blue circles are the input values of the survey parameters, the red circle is the estimated output value)

The results of the simulations are shown in Table 25.

Table 25: Summary of the sample size calculations performed for agricultural and urban areas of the buffer zone

Land use	Design prevalence (%)	Confidence level (%)	Risk level	Relative risk	Convenience sampling	Samples
Agricultural area	0.04	78	High risk	2	2	**2,784**
			Baseline	1	1	**1,392**
Urban area	0.1	78	N/A	N/A	N/A	**2,751**
Total	0.04	87.99				**6,927**
	0.1	99.56				**6,927**

After conducting this annual survey of the buffer zone, it is possible to conclude with 87.99% confidence, that if all 6,927 samples test negative for *X. fastidiosa*, the prevalence of the bacterium, if

present, will be below 0.04 or equivalently with 99.56% confidence, that if all 6,927 samples test negative for *X. fastidiosa*, the prevalence of the bacterium, if present, will be below 0.1%.

Conclusions

At the request of the European Commission, to support the EU Member States, EFSA prepared specific guidelines for the survey of *Xylella fastidiosa*. This document guides the surveyor through the design of statistically sound and risk-based surveys for *X. fastidiosa*, integrating into the design the key information gathered from the pest survey card for *X. fastidiosa* (EFSA, 2019a).

Three different survey aims are distinguished: *detection surveys* to substantiate pest freedom in an area or country, *delimiting surveys* to determine the boundaries of an infested zone, and *buffer zone surveys* to monitor a zone that serves as a buffer around an infested zone and therefore should ensure pest detection at low levels of prevalence. The guidelines have been developed using examples to illustrate the design of these three types of surveys.

The first step of the survey design is to set the aim of the survey, and to characterise the host plant population as well as the identification method for the pest. It will be necessary to quantify the survey parameters and to consider the importance of the assumptions that are made for each one of them. When setting the design prevalence and the confidence level of the survey, the chosen values should reflect the aim of the survey and the compromise between the resources needed to carry out the survey and the risk that risk managers are willing to accept. Good information on land use in the survey area is needed to determine the size of the target population and its hierarchical structure. The host plant population can then be defined by subdividing it into units that are homogeneous in terms of the epidemiology of *X. fastidiosa.* The use of risk factors will allow the surveys to be better targeted to those areas where the probability of infection is higher. The relative risks can be estimated using expert knowledge or by means of data analysis. The method sensitivity needs to be estimated by combining sampling effectiveness and diagnostic sensitivity, which is particularly challenging for *X. fastidiosa* because the method sensitivity varies depending on the host species and a conservative approach is recommended here. The better the information used to establish the survey parameters, the more robust the conclusions of the survey will be.

In the second step, the sample size is calculated using the survey parameters as input for the statistical tool RiBESS+ which calculates the sample size using a statistically sound and risk-based approach. The mathematical principles behind the tool are fully in line with the recommendations and guidelines provided by the different ISPMs. In addition, RiBESS+ is routinely used for surveillance activities in the animal health sector. The approach is further tailored to the surveys of *X. fastidiosa* and illustrated using examples.

The final step of the survey design is the allocation of the samples within subdivisions of the target population. Depending on the information available on the target population and risk factors, the allocation of the samples can be proportional to the number of epidemiological units, or to the size of the host plant population or to the number of risk locations in each region of the survey area. If no information is available, the samples could be allocated at random across the entire survey area.

The robustness of the conclusions of the surveys designed using the proposed approach depends strongly on the quality of the design. The proposed methodology allows one to compare surveys across time and space, thus contributing to harmonisation of surveys in the EU MSs.

Considering that the survey obligations are at EU MS level, and that the data required for survey design are available at national or even regional level, the developed approach should be tailored to each specific situation in terms of host plants, vectors, climate and resources. The approach and tools provided for the specific surveys of *X. fastidiosa* are quite flexible and the success of the design procedure relies on the technical aspects of the survey preparation and the involvement of risk managers.

References

Cameron A, 2012. The consequences of risk-based surveillance: Developing output-based standards for surveillance to demonstrate freedom from disease. Preventive veterinary medicine, 105, 280–286.

Cannon R, 2002. Demonstrating disease freedom - combining confidence levels. Preventive veterinary medicine, 52, 227–249.

EFSA (European Food Safety Authority), in preparation. General guidelines for statistical and risk-based plant pest surveillance.

EFSA (European Food Safety Authority), 2018. Scientific report on the update of the *Xylella* spp. host plant database. EFSA Journal 2018;16(9):5408, 87 pp.

EFSA (European Food Safety Authority), 2020. Scientific report on the update of the *Xylella* spp. host plant database – systematic literature search up to 30 June 2019. EFSA Journal 2020;18(4):6114, 61 pp. doi: 10.2903/j.efsa.2020.6114

EFSA (European Food Safety Authority), Ciubotaru RM, Cortiñas Abrahantes J, Oyedele J, Parnell S, Schrader G, Zancanaro G and Vos S, 2018. Technical report of the methodology and work-plan for developing plant pest survey guidelines. EFSA supporting publication 2018: EN-1399. 36 pp. doi:10.2903/sp.efsa.2018.EN-1399

EFSA (European Food Safety Authority), Vos S, Camilleri M, Diakaki M, Lázaro E, Parnell S, Schenk M, Schrader G, and Vicent A, 2019a. Pest survey card on *Xylella fastidiosa*. EFSA supporting publication 2019:EN-1667. 53 pp. doi:10.2903/sp.efsa.2019.EN-1667

EFSA (European Food Safety Authority), Baker R, Gilioli G, Behring C, Candiani D, Gogin A, Kaluski T, Kinkar M, Mosbach-Schulz O, Neri FM, Preti S, Rosace MC, Siligato R, Stancanelli G and Tramontini S, 2019b. *Xylella fastidiosa* - Pest Report and Datasheet to support ranking of EU candidate priority pests. doi: 10.5281/zenodo.2789977

EFSA PLH Panel (EFSA Panel on Plant Health), 2015. Scientific Opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction options. EFSA Journal 2015;13(1):3989, 262 pp. doi: 10.2903/j.efsa.2015.3989

EFSA PLH Panel (EFSA Panel on Plant Health), Jeger M, Caffier D, Candresse T, Chatzivassiliou E, Dehnen-Schmutz K, Gilioli G, Gregoire J-C, Jaques Miret JA, MacLeod A, Navajas Navarro M, Niere B, Parnell S, Potting R, Rafoss T, Rossi V, Urek G, V an Bruggen A, V an der W erf W, West J, Winter S, Almeida R, Bosco D, Jacques M-A, Landa B, Purcell A, Saponari M, Czwienczek E, Delbianco A, Stancanelli G and Bragard C, 2018. Scientific Opinion on the updated pest categorisation of *Xylella fastidiosa*. EFSA Journal 2018;16(7):5357, 61 pp. doi: 10.2903/j.efsa.2018.5357

EFSA PLH Panel (EFSA Panel on Plant Health), 2019. Update of the Scientific Opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory. EFSA Journal 2019;17(5):5665, 200 pp. doi: 10.2903/j.efsa.2019.5665

EPPO (European and Mediterranean Plant Protection Organisation), 2019. PM 7/24 (4) Diagnostic standard for *Xylella fastidiosa*. EPPO Bulletin, 49 (2), 175–227. doi: 10.1111/epp.12575

Eurostat, online. Database for Total Surface Area by NUTS 3 region. Available online: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=reg_area3&lang=en [Accessed: 4 February 2020]

Eurostat, 2018. Regions in the European Union Nomenclature of territorial units for statistics - NUTS 2016/EU-28 edition 2018. Available online: https://ec.europa.eu/eurostat/documents/3859598/9397402/KS-GQ-18-007-EN-N.pdf/68c4a909-30b0-4a90-8851-eddc400a5faf [Accessed: 12 May 2020]

FAO (Food and Agriculture Organization of the United Nations), 2014. Risk-based disease surveillance – A manual for veterinarians on the design and analysis for demonstration of freedom from disease. FAO Animal Production and Health Manual No. 17. Rome, Italy.

FAO (Food and Agriculture Organization of the United Nations), 2016. ISPM (International Standards for Phytosanitary Measures) 9. Guidelines for pest eradication programmes. FAO, Rome, Italy. Available online: https://www.ippc.int/en/publications/611/

FAO (Food and Agriculture Organization of the United Nations), 2018. ISPM (International Standards for Phytosanitary Measures) 27. Diagnostic protocols for regulated pests. DP (Diagnostic Protocol) 25: *Xylella fastidiosa*. FAO, Rome, Italy. Available online: https://www.ippc.int/en/publications/86498/

FAO (Food and Agriculture Organisation of the United States), 2020. ISPM (International Standard for Phytosanitary Measures) 5. Glossary of phytosanitary terms. FAO, Rome, Italy. Available online: https://www.ippc.int/en/publications/622/

Hornero A, Hernández-Clemente R, North PRJ, Beck PSA, Boscia D, Navas-Cortes JA and Zarco-Tejada PJ, 2020. Monitoring the incidence of *Xylella fastidiosa* infection in olive orchards using ground-based evaluations, airborne imaging spectroscopy and Sentinel-2 time series through 3-D radiative transfer modelling. Remote Sensing of Environment, 236, 111480. doi: 10.1016/j.rse.2019.111480

Martelli FS, Fanti E, Rosati C, Martelli M, Bacci G, Martelli ML and Medico E, 2016. Long-term efficacy of microbiology-driven periodontal laser-assisted therapy. European Journal of Clinical Microbiology and Infectious Diseases, 35(3), 423–431.

Martinetti D and Soubeyrand S, 2019. Identifying lookouts for epidemio-surveillance: application to the emergence of *Xylella fastidiosa* in France. Phytopathology, 109, 265–276. doi: 10.1094/PHYTO-07-18-0237-FI

Mastin AJ, van den Bosch F, Gottwald TR, Alonso-Chavez V and Parnell SR, 2017. A method of determining where to target surveillance efforts in heterogeneous epidemiological systems. PLoS Computational Biology, 13, 1–23. doi: 10.1371/journal.pcbi.1005712

Mastin AJ, van den Bosch F, van den Berg F and Parnell S, 2019. Quantifying the hidden costs of imperfect detection for early detection surveillance. Philosophical Transactions of the Royal Society B, 374, 2018261. doi: 10.1098/rstb.2018.0261

Parnell S, van den Bosch F, Gottwald T and Gilligan CA, 2017. Surveillance to inform control of emerging plant diseases. Annual Review Phytopathology, 55, 591–610.

Purcell AH and Saunders SR, 1999. Fate of Pierce's disease strains of *Xylella fastidiosa* in common riparian plants in California. Plant Disease, 83, 825–830.

White SM, Bullock JM, Hooftman DAP and Chapman DS, 2017. Modelling the spread and control of *Xylella fastidiosa* in the early stages of invasion in Apulia, Italy. Biological Invasions, 19, 1825–1837. doi: 10.1007/s10530-017-1393-5

World Bank, online. Netherlands - Agricultural Land (% of Land Area). Available online: https://tradingeconomics.com/netherlands/agricultural-land-percent-of-land-area-wb-data.html [Accessed: 14 May 2020]

General glossary for pest survey

Term	Definition*
Buffer zone	An area surrounding or adjacent to an area officially delimited for phytosanitary purposes in order to minimise the probability of spread of the target pest into or out of the delimited area, and subject to phytosanitary or other control measures, if appropriate (ISPM 5: FAO, 2020).
Component (of a survey)	A component is a survey entity which can be distinguished based on its target population, the detection method (e.g. visual examination, laboratory testing, trapping) and the inspection unit (e.g. vectors, branches, twigs, leaves, fruits). A pest survey comprises various components. The overall confidence of the survey will result from the combination of the different components.
Confidence	The sensitivity of the survey is a measure of reliability of the survey procedure (Montgomery and Runger, 2010). The term **confidence level** is used in 'Methodologies for sampling of consignments' (ISPM 31: FAO, 2016b).
Delimiting survey	Survey conducted to establish the boundaries of an area considered to be infested by, or free from, a pest (ISPM 5: FAO, 2020).
Design prevalence *analogous to the term **level of detection** used in 'Methodologies for sampling of consignments' (ISPM 31: FAO 2016b)*	It is based on a pre-survey estimate of the likely actual prevalence of the pest in the field (McMaugh, 2005). The survey will be designed in order to obtain at least a positive test result when the prevalence of the disease will be above the defined value of the design prevalence. In 'freedom from pest' approaches, it is not statistically possible to say that a pest is truly absent from a population (except in the rare case that a census of a population can be completed with 100% detection efficiency). Instead, the maximum prevalence that a pest could have reached can be estimated, this is called the 'design prevalence'. That is, if no pest is found in a survey, the true prevalence is estimated to be somewhere between zero and the design prevalence (EFSA, 2018).
Detection survey	Survey conducted in an area to determine whether pests are present (ISPM 5: FAO, 2020).
Diagnostic protocols	Procedures and methods for the detection and identification of regulated pests that are relevant to international trade (ISPM 27: FAO, 2016a).
Epidemiological unit *analogous to the term **lot** used in 'Methodologies for sampling of consignments' (ISPM 31: FAO 2016b*)	A homogeneous area where the interactions between the pest, the host plants and the abiotic and biotic factors and conditions would result in the same epidemiology should the pest be present. The epidemiological units are subdivisions of the target population and reflect the structure of the target population in a geographical area. They are the units of interest to which statistics are applied (e.g. a

	tree, orchard, field, glasshouse, or nursery) (EFSA, 2018).
Expected prevalence	In prevalence estimation approaches, it is the proportion of epidemiological units expected to be infested or infested.
Expert knowledge elicitation	A systematic, documented and reviewable process to retrieve expert judgements from a group of experts in the form of a probability distribution (EFSA, 2014).
Host plant	A host plant is a plant species belonging to the host range on which the pest could find shelter, feed or subsist at least for a period of time.
Host range	Species capable, under natural conditions, of sustaining a specific pest or other organism (ISPM 5: FAO, 2020). This definition is limited to array of host plants species and does not include the commodities other than plants or plant parts.
Identification	Information and guidance on methods that either used alone or in combination lead to the identification of the pest (ISPM 27: FAO, 2016a).
Infected versus infested	Infected is used when a pathogen is referred to in relation to its hosts (e.g. the trees are infected by the bacterium). Infested is used when an insect is referred to in relation to its hosts (e.g. the trees are infested by beetles). Infested is used when the pest is mentioned in relation to an area (e.g. an infested zone).
Inspection	Official visual examination of plants, plant products or other regulated articles to determine whether pests are present or to determine compliance with phytosanitary regulations (ISPM 5: FAO, 2020).
Inspection unit *analogous to* ***sample unit*** *used in 'Methodologies for sampling of consignments' (ISPM 31: FAO 2016b*)	The inspection units are the plants, plant parts, commodities or pest vectors that will be scrutinised to identify and detect the pests. They are the units within the epidemiological units that could potentially host the pests and on which the pest diagnosis takes place (EFSA, 2018).
Inspector	Person authorised by a national plant protection organisation to discharge its functions (ISPM 5: FAO, 2020).
Method sensitivity *analogous to the term* ***efficacy of detection*** *used in 'Methodologies for sampling of consignments'*	The conditional probability of testing positive given that the individual is diseased (Dohoo et al., 2010). The method sensitivity (MeSe) is defined as the probability that a truly positive host tests positive. It has two components: the sampling effectiveness (i.e. probability of selecting infested plant parts from an infested plant) and the diagnostic sensitivity (characterised by the visual inspection

(ISPM 31: FAO 2016b)	and/or laboratory test used in the identification process). The diagnostic sensitivity is the probability that a truly positive epidemiological unit will result positive and is related to the analytical sensitivity. It corresponds to the probability that a truly positive inspection unit or sample will be detected and confirmed as positive. The sampling effectiveness depends on the ability of the inspector to successfully choose the infested plant parts in a host plant. It is directly linked to the sampling procedure itself and on the training of the inspectors to recognise the symptomatology of the pest. Furthermore, symptom expressions are dependent, among other factors, on the weather conditions as well as on the physiological stage of the host plant when the sample is taken.
Pest diagnosis	The process of detection and identification of a pest (ISPM 5: FAO, 2020).
Pest freedom	Pest freedom can be defined, for a given target population, in a statistical framework, as the confidence of freedom from a certain pest against a pre-set design prevalence (threshold of concern).
Population size	The estimation of the number of the plants in the region to be surveyed (EFSA, 2018).
Relative risk	The ratio of the risk of disease in the exposed group to the risk of disease in the non-exposed group (Dohoo et al., 2010).
Representative sample	A sample that describes very well the characteristics of the target population (FAO, 2014).
RiBESS+	Risk-based surveillance systems. This is an online application that implements statistical methods for estimating the sample size, global (and group) sensitivity and probability of freedom from disease. Free access to the software with prior user registration is available at https://shiny-efsa.openanalytics.eu/
Risk assessment	Evaluation of the probability of the introduction and spread of a pest and the magnitude of the associated potential economic consequences (ISPM 5: FAO, 2020).
Risk factor	A factor that may be involved in causing the disease (FAO, 2014). It is defined as a biotic or abiotic factor that increases the probability of infestation of the epidemiological unit by the pest. The risk factors relevant for the surveillance should have more than one level of risk for the target population. For each level, the relative risk needs to be estimated as the relative probability of infestation compared with a baseline with a level 1. Consideration of risk factors in the survey design allows the survey efforts to be enforced in those areas, where the highest probabilities

	exist to find the pest.
Risk-based survey	A survey design that considers the risk factors and enforces the survey efforts in the corresponding proportion of the target population.
SAMPELATOR	Sample size calculator. This is an online application that implements statistical methods to estimate the sample size for pest prevalence estimation surveys. Free access to the software with prior user registration is available at https://shiny-efsa.openanalytics.eu/
Sample size	The sample size refers to the output of the statistical tools for survey design (RiBESS+ and SAMPELATOR). 'A well-chosen sample will contain most of the information about a particular population parameter but the relation between the sample and the population must be such as to allow true inferences to be made about a population from that sample.' (BMJ, online). The survey sample consists of the required number of 'inspection units' or samples thereof to be examined and/or tested in the survey to retrieve sufficient information on the pest presence or prevalence in the total population. In the case of risk-based surveys, the sample size is calculated on the basis of statistical principles that integrate risk factors. If the examination for pest presence is performed by laboratory testing, at least one sample is taken from each inspection unit. These samples will undergo relevant laboratory testing.
Sampling effectiveness	For plants, it is the probability of selecting infested plant parts from an infested plant. For vectors, it is the effectiveness of the method to capture a positive vector when it is present in the survey area. For soil, it is the effectiveness of selecting a soil sample containing the pest when the pest is present in the survey area.
Specified plant	The plant species known to be susceptible to the pest. For example, for *Phyllosticta citricarpa*, the list of specified plants, which includes host plants and all plants for planting, other than seeds, belonging to the genera or species, can be found in Annex I of Decision (EU) 2015/789.
Survey	An official procedure conducted over a defined period of time to determine the characteristics of a pest population or to determine which species are present in an area (ISPM 5: FAO, 2020).
Target population *analogous to* ***consignment*** *used in 'Methodologies for*	The set of individual plants or commodities or vectors in which the pest under scrutiny can be detected directly (e.g. looking for the pest) or indirectly (e.g. looking for symptoms suggesting the presence of the pest) in a given habitat or area of interest. The different components pertaining to the target population that need to be specified are:

sampling of consignments' (ISPM 31: FAO 2016b)	• definition of the target population: the target population has to be clearly identified; • target population size and geographic boundary. (EFSA, 2018)
Test	Official examination of plants, plant products or other regulated articles, other than visual, to determine whether pests are present, identify pests or determine compliance with specific phytosanitary requirements (ISPM 5: FAO, 2020).
Test specificity	The conditional probability of testing negative given that the individual does not have the disease of interest (Dohoo et al., 2010). The test diagnostic specificity is the probability that a truly negative epidemiological unit will give a negative result and is related to the analytical specificity. In freedom from disease it is assumed to be 100%.
Visual examination	The physical examination of plants, plant products, or other regulated articles using the unaided eye, lens, stereoscope or microscope to detect pests or contaminants without testing or processing (ISPM 5: FAO, 2020).

*References

BMJ (British Medical Journal), online. Populations and samples. Available online: https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/3-populations-and-samples [Accessed: 13 March 2020]

Dohoo I, Martin W and Stryhn H, 2010. Veterinary epidemiologic research. 2nd Edition. VER Inc., Canada.

EFSA (European Food Safety Authority), 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734. [278 pp.] doi: 10.2903/j.efsa.2014.3734 Available online: www.efsa.europa.eu/efsajournal

EFSA (European Food Safety Authority), 2018. Technical report of the methodology and work-plan for developing plant pest survey guidelines. EFSA supporting publication 2018: EN-1399. 36 pp. doi:10.2903/sp.efsa.2018.EN-1399

FAO (Food and Agriculture Organization of the United Nations), 2014. Risk-based disease surveillance – A manual for veterinarians on the design and analysis for demonstration of freedom from disease. FAO Animal Production and Health Manual No. 17. Rome, Italy.

FAO (Food and Agriculture Organization of the United Nations), 2016a. ISPM (International Standards for Phytosanitary Measures) 27. Diagnostic protocols for regulated pests. FAO, Rome, Italy. Available online: https://www.ippc.int/en/publications/593/

FAO (Food and Agriculture Organization of the United Nations), 2016b. ISPM (International Standards for Phytosanitary Measures) 31. Methodologies for sampling of consignments. FAO, Rome, Italy. Available online: https://www.ippc.int/en/publications/588/

FAO (Food and Agriculture Organisation of the United States), 2020. ISPM (International Standard for Phytosanitary Measures) 5. Glossary of phytosanitary terms. FAO, Rome, Italy. Available online: https://www.ippc.int/en/publications/622/

McMaugh T, 2005. Guidelines for surveillance for plant pests in Asia and the Pacific. ACIAR Monograph No.119, 192 pp.

Montgomery DC and Runger GC, 2010. Applied statistics and probability for engineers. Fifth Edition, John Wiley & Sons. 792 pp.

Abbreviations

EFSA	European Food Safety Authority
EPPO	European and Mediterranean Plant Protection Organisation
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
IPPC	International Plant Protection Convention
ISPM	International Standards for Phytosanitary Measures
MS(s)	Member State(s)
NL	Netherlands
NPPO	National Plant Protection Organisation
NUTS	Nomenclature of Territorial Units for Statistics
PACA	Provence Alpes Côte d'Azur
PCR	Polymerase chain reaction
PIZ	Potentially infested zone

Appendix A – Using Corine Classes to identify different environments for survey design

The Corine Land Cover database is available at https://land.copernicus.eu/pan-european/corine-land-cover/clc2018

Land classes can be extracted for any spatial area and then further classified as epidemiologically relevant. In the example, we extracted data for the *Xylella fastidiosa* demarcated area in Valencia (Table A2) and collated these in four distinct environments: forest, semi-wild, agricultural and urban (Figure A1; Table A1). Land cover classes that have no epidemiological significance for the pest have been omitted.

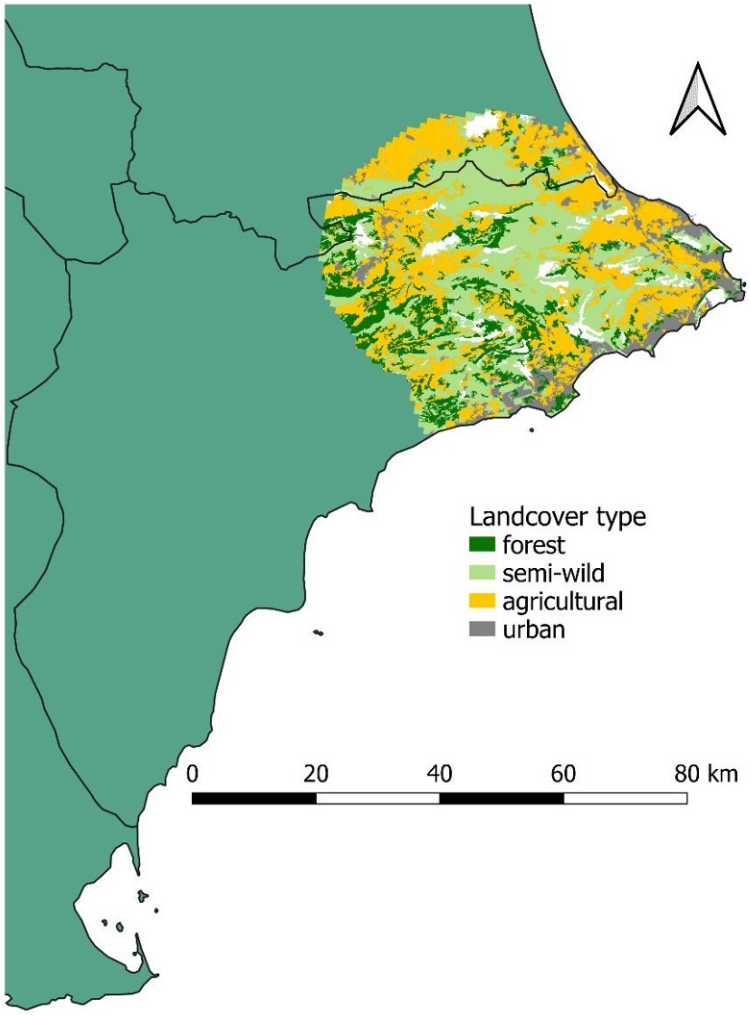


Figure A1: Environment classes for the *Xylella* demarcated area in Valencia, Spain

Table A1: Land use categories within the demarcated area in Valencia

Land use category	**Area (ha.)**
omitted	14,293.5
semi-wild	83,737.48
urban	21,341.74
agricultural	79,778.02
forest	32,328.55

Table A2: Corine landcover classes (CLC) in the *Xylella fastidiosa* demarcated area in Valencia and their new classification by *Xylella fastidiosa* (Xf) environment type

CLC CODE	**Description**	**Xf environment**
111	Continuous urban fabric	urban
112	Discontinuous urban fabric	urban
121	Industrial or commercial units	urban
122	Road and rail networks and associated land	urban
123	Port areas	urban
124	Airports	urban
131	Mineral extraction sites	semi-wild
132	Dump sites	urban
133	Construction sites	urban
141	Green urban areas	semi-wild
142	Sport and leisure facilities	urban
211	Non-irrigated arable land	agric
212	Permanently irrigated land	agric
213	Rice fields	agric
221	Vineyards	agric
222	Fruit trees and berry plantations	agric
223	Olive groves	agric
231	Pastures	agric
241	Annual crops associated with permanent crops	agric
242	Complex cultivation patterns	agric
243	Land principally occupied by agriculture, with significant areas of natural vegetation	agric
244	Agro-forestry areas	forest
311	Broad-leaved forest	forest
312	Coniferous forest	forest
313	Mixed forest	forest
321	Natural grasslands	semi-wild
322	Moors and heathland	semi-wild
323	Sclerophyllous vegetation	semi-wild
324	Transitional woodland-shrub	semi-wild
331	Beaches, dunes, sands	omit
332	Bare rocks	omit
333	Sparsely vegetated areas	omit
334	Burnt areas	omit
335	Glaciers and perpetual snow	omit
411	Inland marshes	omit
412	Peat bogs	omit
421	Salt marshes	omit
422	Salines	omit
423	Intertidal flats	omit
511	Watercourses	omit
512	Water bodies	omit
521	Coastal lagoons	omit

522	Estuaries	omit
523	Sea and ocean	omit
999	No data	omit
990	Unclassified land surface	omit
995	Unclassified water bodies	omit

Appendix B – Probability of *Xylella fastidiosa* infection of host plants genus

As mentioned in Section 1.2.2, *Xylella fastidiosa* is able to infect a wide range of potential host species. This fact together with the variability of diagnostic method sensitivity for the various host plants, as well as the differences in host plant susceptibility, presents a significant challenge for the design of any type of survey.

Based on the data collected in the current European outbreak zones, a battery of statistical analyses were carried out to extract information on the differences between host plant genera susceptibility to be used in the various types of *X. fastidiosa* survey designs. The data were kindly provided to EFSA by the corresponding authorities for the EU outbreaks in Spain (Alicante (2017–2018) and the Balearic Islands (2017–2018)), France (Provence Alpes Côtes d'Azur (PACA) and Corsica (2015–2018)) and Italy (Apulia (2013–2019)) and contain information on 467,635 host plant samples belonging to 298 genera and 467 species. The vast majority of them were characterised by their corresponding (i) genus, (ii) species, (iii) date of collection, (iv) *X. fastidiosa*-like symptoms (presence or absence), (v) *X. fastidiosa* presence (detected or undetected), (v) geolocation, and (vi) outbreak origin. However, as only a small proportion was described at the pest subspecies or sequence type level that information was not used.

Before the analyses, different datasets were generated by the combination of some of the following inclusion criteria (see Table B1):

a) Exposure to the bacterium. A cluster algorithm was implemented to build hypothetical 'infested areas' based on the information provided by the sample geolocation and the date of collection. Thus, only samples which were located within these exposure areas were selected.

b) Sufficient sample size and ability to become infested. Only genera with more than 100 samples and a minimum of one positive sample were selected.

c) Ability to express symptoms. Only genera in which *X. fastidiosa*-like symptoms were observed were considered.

They were subsequently analysed at genus level to estimate:

1. the probability of infection of the most representative genera;
2. the probability of infection of the most representative genera on a regional scale;
3. the probability of infection of the most representative genera given the presence of symptoms on a regional scale.

The most representative genera were selected after this analysis in order to select only those categories which presented the most accurate estimates.

Table B1 shows the probability of infection estimated according to the survey strategy to be used and the scope of application within the EU territory. Note that the information provided is based on the information available on the current outbreaks and it should be updated with new findings of the bacterium and updates on the current ones.

Table B1: Estimated probabilities, survey strategy and scope of application within EU territory

Estimated probabilities	Survey strategy	Scope of application
Probability of the most representative genera becoming infected with *Xylella fastidiosa* (*) (Table B2)	Detection and buffer zone surveys	All MSs
Probability of the most representative genera for each outbreak region becoming infected with *Xylella fastidiosa* (**) (Tables B3 and B4)	Detection surveys and buffer zone surveys	MSs affected
Probability of the most representative genera for each outbreak region becoming infected by *Xylella fastidiosa* given the presence of symptoms (***) (Table B5 and B6)	Delimiting surveys	MSs affected

(*) Estimated using a data subset generated considering inclusion criteria (a) and (b).
(**) Estimated using a data subset generated considering inclusion criteria (a) and (b) and filtering by outbreak origin.
(***) Estimated using a data subset generated considering inclusion criterion (c) and filtering by outbreak origin.

Table B2: Estimated probability of the most representative genera becoming infected by *Xylella fastidiosa* according to the current EU *Xylella* outbreak information

Detection and buffer zone surveys	
Genera	**Probability of infection**
Polygala	0.551
Helichrysum	0.511
Euryops	0.471
Calicotome	0.452
Genista	0.315
Spartium	0.161
Lavandula	0.152
Cistus	0.126
Prunus	0.093
Olea	0.076
Vitis	0.057

Table B3: Estimated probability of the most representative genera becoming infected by *Xylella fastidiosa* according to the current Alicante and Balearic Islands outbreak information

Detection and buffer zone surveys			
Spain (Alicante)		**Spain (Balearic Islands)**	
Genus	**Probability of infection**	**Genus**	**Probability of infection**
Polygala	0.5	*Rhamnus*	0.591
Prunus	0.489	*Prunus*	0.495
		Olea	0.448
		Polygala	0.381
		Vitis	0.276
		Rosmarinus	0.219
		Ficus	0.123

Table B4: Estimated probability of the most representative genera becoming infected by *Xylella fastidiosa* according to the current Corsica–PACA and Apulia outbreak available information

Detection and buffer zone surveys			
France (Corsica and PACA)		**Italy (Apulia)**	
Genus	**Probability of infection**	**Genus**	**Probability of infection**
Polygala	0.588	*Olea*	0.076
Helichrysum	0.561	*Nerium*	0.010
Calicotome	0.518	*Prunus*	0.002
Euryops	0.160		
Spartium	0.179		
Genista	0.435		
Lavandula	0.471		
Cistus	0.315		
Pelargonium	0.111		

Table B5: Estimated probability of the most representative genera becoming infected by *Xylella fastidiosa* given the presence of symptoms according to the current Alicante and Balearic Islands outbreak information

Delimiting surveys			
Spain (Alicante)		**Spain (Balearic Islands)**	
Genus	**Probability of infection**	**Genus**	**Probability of infection**
Prunus	0.341(*)	*Prunus*	0.395
		Olea	0.295
		Polygala	0.281
		Vitis	0.106

(*) With the available data, there was not enough information to obtain robust estimates for the other sampled genera.

Table B6: Estimated probability of the most representative genera becoming infected by *Xylella fastidiosa* given the presence of symptoms according to the current Corsica–PACA and Apulia outbreak information

Delimiting surveys			
France (Corsica and PACA)		**Italy (Apulia)**	
Genus	**Probability of infection**	**Genus**	**Probability of infection**
Calicotome	0.964	*Olea*	0.220(*)
Phagnalon	0.900		
Helichrysum	0.836		
Genista	0.578		
Polygala	0.536		
Cistus	0.491		
Pelargonium	0.472		
Spartium	0.298		
Lavandula	0.297		
Euryops	0.270		

(*) With the available data, there was not enough information to obtain robust estimates for the other sampled genera.

The information provided in Tables B2–B6 can be used to inform the selection of host plants, integrating the information into the inspection and sampling instructions. Furthermore, it can also be further exploited as a risk factor to better allocate survey efforts. Note that to manage this information as a risk factor, after the selection of the genus of interest, each selected category must be characterised by its relative risk and its proportion within the target population. Relative risk computation can be calculated as a ratio of the probability of infection of genus A versus genus B with genus B as the reference genus (i.e. the genus with the smallest probability), thus relative risk can vary depending on the choice of that reference genus and it can be adapted in each survey design according the list of genera selected to better characterise the survey area (see Section 2.4.2 for further details).

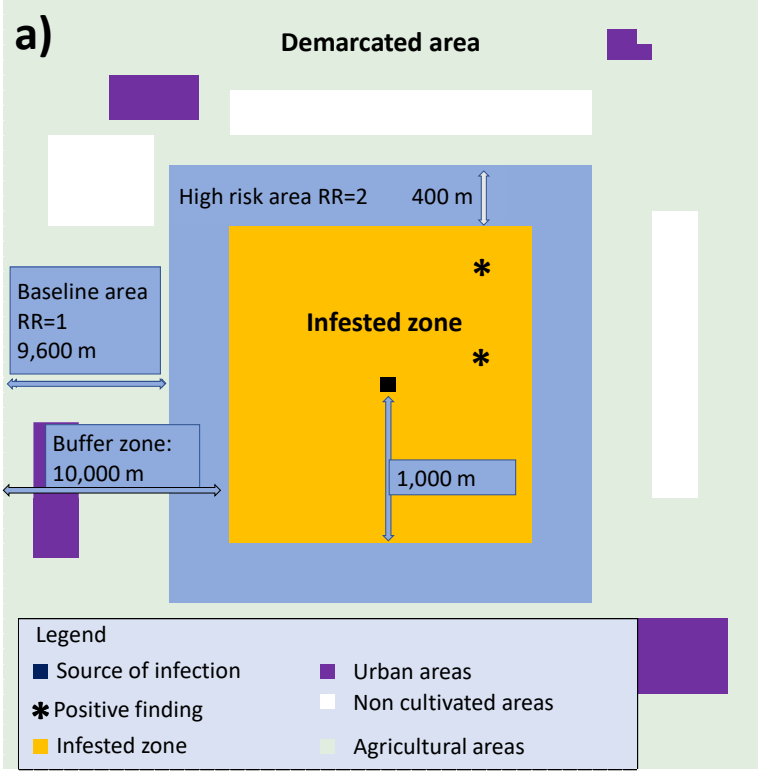
Appendix C – Design of a buffer zone survey for *Xylella fastidiosa* in a two-step approach

C.1. Survey design scenario

Following the example given in Section 4, an infected plant has been found in an agricultural area and the area has been demarcated with an infested zone and a buffer zone as shown in Figure C1.

The scenario of the simulations in this appendix is the annual survey performed in the agricultural areas of the buffer zone to verify its freedom from *Xylella fastidiosa*. The other land use categories could be addressed the same way.

a)



b)

Area of Buffer zone = Area of Demarcated area - Area of infested zone

$(10{,}000+10{,}000+1{,}000+1{,}000+1)^2 - (1{,}000+1{,}000+1)^2$

= 48,004 hectares

c)

	Hectares	Host plants/ha	Host plant population
Agricultural area	25,000	300	7,500,000
Urban area	12,000	100	1,200,000
Inaccessible areas or areas without host plant (e.g. rocky areas, roads)	11,000	N/A	N/A
Buffer zone	**48,000**		**9,300,000**

Figure C1: Characteristics of the buffer zone: (a) schematic representation; (b) calculation of the area covered by the buffer zone; (c) land use categories and areas in the buffer zone

C.2. Survey parameters

C.2.1. Target population and epidemiological units

As defined in the Glossary, an epidemiological unit is a homogeneous area where the interactions between the pest, the host plants, the abiotic and biotic factors and conditions would result in a similar epidemiology if the pest was present. The epidemiological units are subdivisions of the target population according to an epidemiological homogeneity criterion and reflect the structure of the target population in a geographical area. They are the units of interest for which the sample size is estimated. This could be achieved by calculating the overall sample size and then proportionally allocating them to each subset of the target population. For a statistically based survey it is therefore essential to clearly define these epidemiological units, indicating the related assumptions.

To optimise the survey efforts in terms of the number of samples that represent the host population, as much information as possible should be gathered on the homogeneity of the survey area and epidemiological units should be chosen for which the homogeneity assumptions are realistic and acceptable. The homogeneity should be analysed in terms of ecology (habitat, environmental suitability, timing of life stages in the year, crops, host plants, vector abundance, etc.), exposure (pathways and entry points, flora, etc.), geographical and topographical characteristics.

When there is little information on the epidemiological homogeneity available for the whole survey area or for each land use category, an extreme case would be to consider each hectare that contains at least one host plant of *X. fastidiosa* as an independent epidemiological unit. In this case the assumption taken on homogeneity is likely to be fulfilled. As a consequence of the high number of epidemiological units, the resulting sample size would also be very high. However, despite the high number of samples, this approach enables a practical and simple allocation of the samples to be taken in the survey.

C.2.2. Confidence level and design prevalence

The survey has been set up to achieve a 95% confidence level with a design prevalence of 0.4% in the agricultural area.

In an orchard with 300 host trees, if all samples are tested negative, it would be possible to be 95% confident that, if the pest is circulating in the orchard, it would infect less than 0.4% of the trees.

C.2.3. Method sensitivity

The method sensitivity reflects how good the method is at detecting the pest when it is present. Method sensitivity combines sampling effectiveness and diagnostic sensitivity values. 0.55 has been used as the reference value for the simulations. Further explanations are provided in Section 2.3.

C.2.4. Summary of survey parameters

The information required for the use of the statistical tool RiBESS+ is summarised in Table C1.

Table C1: Summary of the survey parameters for a survey of the agricultural areas of the buffer zone

Survey parameters		
Target population	Agricultural area 300 plants/ha	7,500,000 host plants
	Epidemiological unit	1 ha
	Inspection unit	1 host plant
Risk factors	High-risk area 400 m band adjacent to infested zone	400 ha 120,000 host plants Relative risk 2
	Baseline area	24,600 hectares 7,380,000 host plants Relative risk 1
Design prevalence		0.4%
Confidence level		95%
Method sensitivity		0.55

C.3. Two-step approach

The survey design is developed by first estimating the number of host plants that need to be sampled within each single hectare (Step 1) and then by estimating the number of hectares that need to be inspected (Step 2) to achieve a 95% confidence level for detecting the pest above 0.4% prevalence.

The confidence level that is achieved when calculating the number of host plants to sample within each hectare in the first step becomes the sensitivity of the method for calculating the number of hectares that need to be inspected in the second step.

As a consequence, the higher the confidence within the hectare, the fewer hectares need to be inspected, and inversely if the confidence at hectare level is decreased, the more hectares need to be inspected to achieve an overall 95% confidence in the survey.

The resulting sample size is calculated by multiplying the number of hectares that need to be surveyed by the sample size calculated within the single hectare.

Step 1: Number of samples within each hectare

In this first step, the objective is to calculate the number of plants that need to be sampled out of 300 plants with a design prevalence of 0.4% and a method sensitivity of 0.55.

The confidence level at hectare level needs to be set. This value depends on the number of host plants in each hectare and on the method sensitivity (Table C2). The lower these two values are, the lower the confidence within the hectare will have to be. This is because the number of host plants might not be sufficient to achieve a high level of confidence when the method sensitivity is low (red cells in Table C2).

When selecting the confidence at hectare level, practical reasons could also be considered as it might be more convenient to inspect more plants per hectare but visit fewer hectares.

For the simulations, the scenario was selected that had a 20% confidence level at hectare level and 93 host plants per hectare to sample (Table C2, green cells, and Figure C2).

The choice of the host plants to sample can be prioritised according to the probability of the host plant genus becoming infected as determined in Appendix B. In this exercise this risk factor is integrated into the sampling procedure and will not be used in the calculation of the sample size.

Table C2: Confidence level and number of samples within each hectare calculated with RiBESS+, for a survey at hectare level with the following survey parameters: host population: 300 plants; design prevalence: 0.4%; method sensitivity: 0.55

Confidence level	10%	**20%**	30%	40%	50%	60%	65%
Sample size per hectare	46	**93**	141	190	240	292	Impossible as there are more samples to take than host plants in a hectare

Figure C2 shows how to use RiBESS+ to calculate the sample size per hectare to achieve a 20% confidence level at the hectare level.

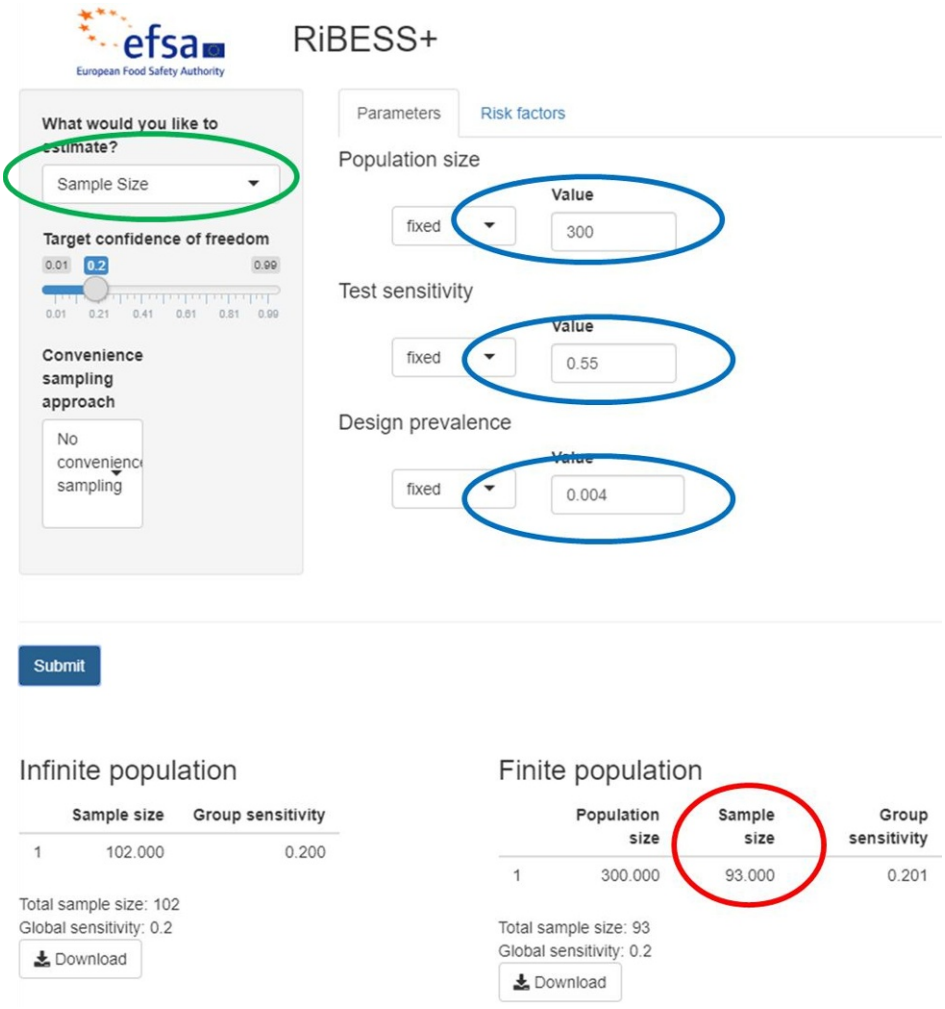


Figure C2: Screenshot of RiBESS+ calculating the sample size per hectare to achieve 20% confidence and 0.4% design prevalence with a method sensitivity of 55% and 300 host plants per hectare (the green circle is the chosen functionality, the blue circles are the input values of the survey parameters, the red circle is the estimated output value)

Step 2: number of hectares to inspect

In this second step, the objective is to calculate the number of hectares to inspect when 93 trees are sampled in each inspection, to achieve an overall confidence of 95% of detecting 0.4% of infections.

The sensitivity of the method to indicate that a hectare is infested with *X. fastidiosa* is needed to calculate the number of hectares to visit. The sensitivity of the method is, in this case, the probability of assessing a hectare as infested when the hectare is truly infested. In the previous step we defined a hectare as infested when we can be 20% confident that the number of infected host plants is below 0.4%. This confidence level at hectare level becomes here the sensitivity of the method.

In this step, a risk factor is introduced. The host plants in the first band of 400 m adjacent to the infested zone (i.e. the high-risk area) are estimated to be twice as likely to be infected with *X. fastidiosa* as the other host plants (i.e. in the baseline area). This is summarised in Table C1:

- 400 ha of the agricultural area of the buffer zone are the high-risk area
- 24,600 ha of the agricultural area of the buffer zone are the baseline area.

Figure C3 shows how to use RiBESS+ to calculate the number of hectares to inspect with integration of the risk factor to achieve an overall 95% confidence level and a 0.4% design prevalence given a 20% confidence level at the hectare level.

Figure C3: Screenshots of RiBESS+ for calculating the number of hectares to inspect with integration of the risk factor to achieve an overall 95% confidence level and a 0.4% design prevalence given 20% method sensitivity (20% confidence level at the hectare level) (the green circles are the chosen functionalities, the blue circles are the input values of the survey parameters, the red circles are the estimated output values)

Conclusion: number of samples to take in the agricultural area of the buffer zone

The results and combinations of the different calculations for Step 1 and Step 2 are presented in Table C3.

Table C3: Results of the sample size calculation using a two-step approach for a risk-based survey in the agricultural area of the buffer zone (epidemiological units of 1 ha)

Survey steps	Design prevalence	Confidence level	Method sensitivity	Risk factor			Number of samples	
				Risk level	Relative risk	Proportion of hectares	without risk factor	with risk factor
Step 1	0.4%	20%	55%		-	-	93 trees	93 trees
Step 2	0.4%	95%	20%	High risk	2	0.016		400 ha
				Baseline	1	0.984		2,878 ha
				Total hectares			3,688 ha	3,278 ha
Agricultural area of the buffer zone	**0.4%**	**95%**	**-**				**342,984 samples**	**304,854 samples**

The estimated sample sizes result from the extreme scenario where little information is known on the homogeneity of the agricultural area that has been subdivided into single hectares which are considered homogeneous in terms of the epidemiology of *X. fastidiosa*.