RISK-BASED SAMPLING (RBS) MANUAL – PART II

Multi-authored manual

2022

CONTENT

Recording inspection data.

1. INTRODUCTION

***Robert Griffin*[1]**

1. National Coordinator for Agriculture Quarantine Inspection USDA, APHIS, PPQ – Retired

All National plant protection organizations (NPPOs) share the same objectives regarding their international framework of obligations as outlined in the World Trade Organization's Agreement on the Application of Sanitary and Phytosanitary Measures (WTO-SPS) and the International Plant Protection Convention (IPPC): free, fair, and safe trade. Inspection plays a central role in meeting these objectives but practicing inspection in a risk-based way is only beginning to be understood and accepted. One reason for this is the historical focus on inspection as the primary strategy for **pest exclusion** with little emphasis on its role as a **source of information** to monitor pest risk and facilitate analysis that can support risk-based strategies. It is clear from Part I of this manual (Click here) that risk-based sampling (RBS) provides the means to maximize the effectiveness of inspection for pest exclusion while also promoting a fair and technically defendable trade environment. In addition, RBS has the great benefit of providing powerful data for analyses that help NPPOs understand the relationship of risk to resources and the opportunities for adjustments that can improve their approach to risk management.

The best RBS design begins with a good understanding of the individual operational situation where it will be applied. Statistical and analytical tools are then used by the NPPO to determine the best RBS design to meet the desired scope, available resources, and expected outcomes for that country. There is a growing store of experiences and information that can shed light on many aspects of RBS beyond its basic principles. Part II of this manual is a collection of tools and information that draw from the experience of NPPOs and experts to provide additional detail on RBS inspection designs. This compilation of resources offers another level of insight with information, interpretations, and recommendations that may be helpful to those searching for greater depth in their understanding of RBS.

The process of building RBS capacity may be as simple or complex as the NPPO is comfortable undertaking. Although the fundamentals of RBS are relatively simple, their application in practice can be challenging. Risk-based sampling requires some baseline understanding of statistics consistent with all scientific endeavors, including the discipline of risk management. A substantial portion of the RBS manual Part II is devoted to statistical elements that go beyond what is covered in the RBS manual Part I or in the international standards for phytosanitary measures (ISPMs) that provide inspection guidance. Some of this information can be directly useful to those who are fluent in statistics. NPPOs are encouraged to take advantage of the resources provided in Part II but also seek and build statistical capacity as needed to address their unique, individual RBS implementation challenges.

Phytosanitary inspection of Hass avocados from Peru.
Source - https://www.senasa.gob.pe/senasacontigo/ica-inspeccion-fitosanitaria-de-palta-hass-para-exportacion-a-china/

Phytosanitary inspection of chrysanthemum flower bunches from Colombia.
Source - https://www.ica.gov.co/noticias/ica-exporta-pompon-hacia-chile-certificados

2. SPECIAL TOPICS

Robert Griffin[1]

1. National Coordinator for Agriculture Quarantine Inspection USDA, APHIS, PPQ – Retired

Every country has different experiences with the implementation of RBS. The range of experiences to date have helped identify aspects of implementation that deserve special attention. The discussions that follow, aim to provide insight into practical issues that have drawn special attention for the challenges they pose.

2.1 Political commitment

Chapter 5, section 5.1 of the RBS manual Part I, states that "The first prerequisite for successful implementation of RBS is the combination of training and commitment to ensure that the concepts are understood and supported". It is important to reemphasize this point and add that a crucial aspect of this commitment is strong, consistent support from NPPO leadership. The first reaction to suggested change in inspection approaches/designs by the responsible workforce (inspectors) is often resistance to change, especially if it seems counterintuitive on the surface. Stakeholders may also resist this change as they experience adjustments in rates of regulatory actions and in resource allocation. Additionally, the shift from long-practiced inspection methods and beliefs is likely to result in challenges, mistakes, and frustrations that make it tempting to reverse course. Consistent encouragement and demonstrated determination by NPPO leadership is necessary to stay the course. The final goal in the journey towards implementing RBS should be to achieve the "tipping point" where experience leads to understanding its value, embracing it for its fairness and efficiency, and broadly accepting it as the best practice for technically justified risk management.

The final goal in the journey towards implementing RBS should be to achieve the "tipping point" where experience leads to understanding its value, embracing it for its fairness and efficiency, and broadly accepting it as the best practice for technically justified risk management.

2.2 Training

Inspection is an acquired skill. As such, it requires training to **understand its concepts** as a first step and then **practice applying these concepts** as a second step. The best inspection training goes beyond sharing information including hands-on exercises and real/realistic situations that demonstrate the concepts using a variety of circumstances and data, to showing not only the

benefits of RBS, but also its limitations. Finally, the consistent implementation of well-designed RBS programs is necessary to internalize best practices.

2.3 Space and equipment

Adequate time, lighting, space, and equipment are all factors that contribute to good inspection but for practical reasons may not be optimal at every NPPO. Recognizing that some shortcomings exist in almost every situation, inspection designs should take account of the effects of these shortcomings. For instance, pest interceptions may be expected to increase when better lighting, equipment, and inspection space is provided. Likewise, ensuring that inspectors have sufficient time for inspection results in better outcomes.

It is often the case that there may be no change in the risk associated with the cargo, only a change in the conditions for pest detection that results in more or less pest interceptions. Whenever the question is raised about whether risk has changed, an evaluation of the conditions needs to be first considered to understand if the change is related to operational factors before any policy changes are implemented.

2.4 Randomization

A common limitation of inspection is sufficient time, equipment, and a secure space for unloading cargo to fully randomize the consignment before sampling. Random sampling is important from an operational standpoint for discovering differences in risk that may be associated with the cargo configuration. If samples are always taken from the rear of the container, the inspector is unable to gain insights into the characteristics of the cargo in other areas of the container. This practice assumes homogeneity of the cargo but needs to be tested occasionally to verify. Likewise, randomization increases statistical confidence in the results of inspection. An alternative to random sampling every consignment is to randomize some subset of consignments that is practical (e.g., one out of every 30 consignments) and then compare the inspection results to those obtained from non-random sampling to understand the variation in confidence. It is sometimes possible to take advantage of cargo devanning (=unloading) that might be required by Customs or other border agencies to gain access to cargo that is not normally available for inspection. This argues for a high level of collaboration with other border agencies to coordinate inspection designs that take maximum advantage of opportunities to gather better information.

Random sampling is important from an operational standpoint for discovering differences in risk that may be associated with the cargo configuration.

2.5 Selecting the appropriate sample unit

Inspecting apple fruit in a consignment will not help with detecting pests present in the consignment's wood packaging, or snails that may be attached to the outside of the container used to ship the apples. Every inspection needs to include an element of general examination for the detection of unexpected pests to complement the inspection designs for specific pests. Furthermore, the inspection designs for specific pests need to account for the type of pest and its behavior. For instance, pests that feed internally will require destructive sampling of individual fruit (cutting). However, the most appropriate sample unit for mobile pests, external feeders, and contaminants in fruits and vegetables is more likely to be individual packages (e.g., box, bag, tray, etc.). Bulk commodities like grain are usually sampled based on increments of weight or volume (e.g., kilograms, pounds, ounces, etc.). Selection of the appropriate sample unit requires thoughtful consideration and consistent application to ensure meaningful inspection results. The results of inspections based on boxes cannot be easily compared or combined with inspection results based on weight.

2.6 Uncertainty

Uncertainty includes both natural variability and error, and inspection has a mix of both types of uncertainty. Variability may be controlled up to a point but not eliminated. Errors can be corrected where they are discovered. Because inspection is a human activity, there will always be natural variability in the process and therefore in the results, and there will always be some level of error - both factors affect the efficacy of inspection. The few studies that exist on inspection efficacy show a wide range of results, varying from around 20% up to 80% depending on many factors (Gould, 1995). Interestingly, inspectors typically believe that inspection is highly effective but base this perception on their bias for detection where pests have been found before. For this reason, new inspectors are often the ones that find new pests. Risk-based sampling designates inspection units without bias and requires the entire sample to be inspected. Both of these procedures increase the probability of detecting previously undetected pests and support better analysis by providing information on pest prevalence (number of pests in a sample).

Aside from the uncertainty associated with efficacy, there is uncertainty associated with other statistical parameters of inspection. One obvious area of uncertainty is the confidence level. The statistical convention for confidence is 95%. If applied consistently, this means that 5 times out of 100, the results will be wrong. This is in addition to the uncertainty that comes from the tolerance we accept with sampling. If sampling is designed to detect a 10% infestation level with 95% confidence, then we not only have the possibility of 5% incorrect results, but we also have the uncertainty associated with the other 95% where it falls below the detection level of 10%.

The 10% may or may not be infested, but we have decided to live with the possibility that as much as 10% of pests are undetected because they fall under our tolerance threshold.

In sum, there are many areas of uncertainty and the magnitude of uncertainty associated with inspection can be large and vary widely. For this reason, inspection must be designed and used carefully as a phytosanitary measure, recognizing that it provides limited precision for risk mitigation but can provide valuable information for improved overall risk management.

2.7 Correlation of inspection variables

The focus of analysis for RBS data will typically be the regulatory action rate for specific commodities from specific countries. This simple correlation of regulatory actions to commodity/country requires a limited data set with enough observations to be statistically significant (see also 2.8 below). Collecting additional data on other inspection variables greatly expands the possibilities for analysis, and such data is normally collected/available for imports. For instance, the regulatory action rate for a particular country/commodity may be related to a single producer in the country of origin. By collecting producer data, this relationship can be detected and more easily corrected. Likewise, it may be that regulatory actions on the same commodity from the same origin are different when processed at different ports or by different inspectors, or that these regulatory actions are more frequent at certain times of the year. A broad range of useful analyses can be done by correlating different inspection variables using RBS data because it is collected in a consistent way. The key is collecting data on inspection variables that serve a useful purpose for managing risk and not wasting time and resources collecting data that is not helpful for answering the questions that are important for risk and resource management.

2.8 Singling, mingling, and commingling cargo

Cargo configurations can include a mix of products/commodities, different packaging, and different configurations which can create challenges for RBS sampling. Since all possibilities cannot be anticipated for inspection designs, it is important to allow inspectors some flexibility to address these real-world challenges while also maintaining a high level of consistency during sampling.

The most common cargo configuration is known as **singling**. A singled consignment is one where the commodity/product is homogeneous; all the product and packages are the same (**Figure 1**). The main challenge with singled consignments is identifying the sample unit when products have multiple packages (e.g., bags within boxes) or unusual packing that complicates sampling (e.g., trays, towers, bulk stacks). The key to sampling singled consignments is identifying the most appropriate and practical sample unit and then using it consistently.

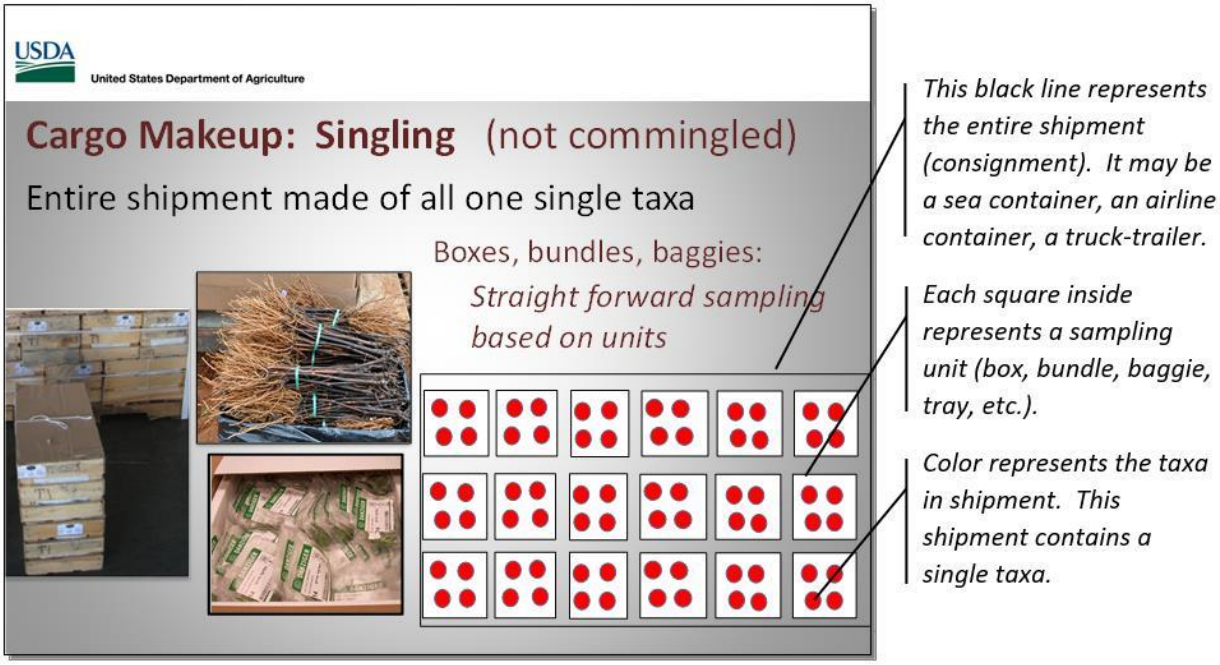


Figure 1. *Singled shipment which is comprised of a single taxon; the red dots represent the taxa in the shipment.*
Source: https://nappo.org/application/files/5415/8676/4129/RBS_Symposium_Proceedings_-10062018-e.pdf

Mingled consignment configurations represent the next step in complexity. A mingled consignment is one where the product/commodity is the same within sample units but there is a mixture of products/commodities in the consignment (**Figure 2**). Imagine a consignment of cut flowers that contains both rose and carnation flowers. The primary challenge with mingled consignments is deciding whether to sample it as a single lot or divide the consignment according to the products and assign different sampling regimes to each. This decision will depend on the products/commodities in question and whether there is an expectation that each presents different risks. In the absence of experience or other information, both can be sampled together as they would be for a singled consignment until a difference is noted that justifies distinguishing one from the other. Another reason to sample differently may be concerns for a specific pest that is known to be associated with one product (rose flowers) and not the other.

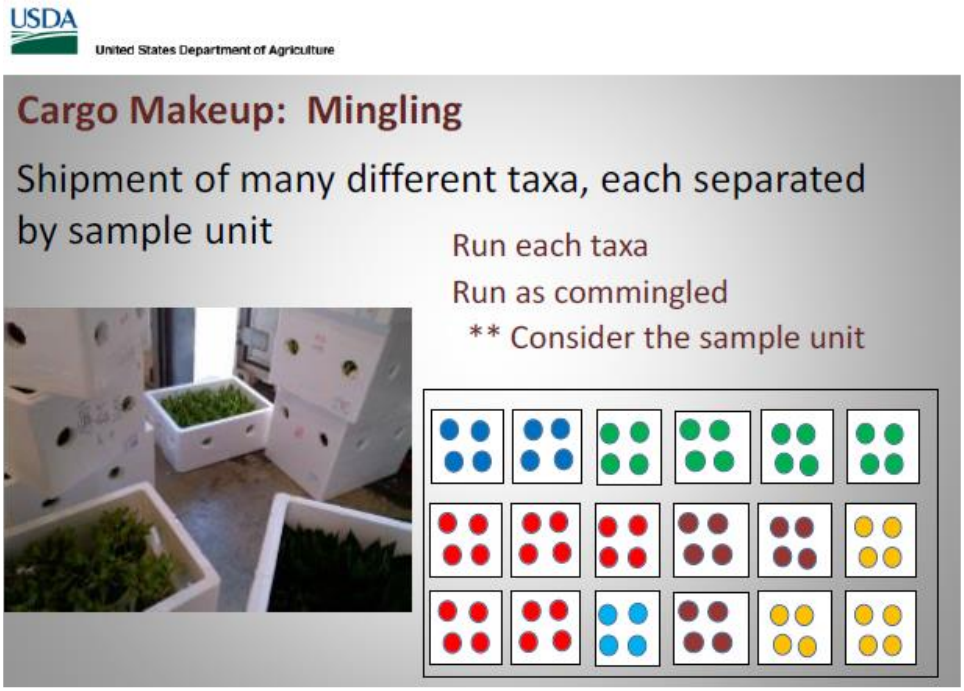


Figure 2. *Mingled shipment; each colored dot represents a different taxon in the shipment*. Source: https://nappo.org/application/files/5415/8676/4129/RBS_Symposium_Proceedings_-10062018-e.pdf

Finally, a **comingled** consignment presents the most challenging situation. Comingled consignments have multiple products in each sample unit (**Figure 3**).

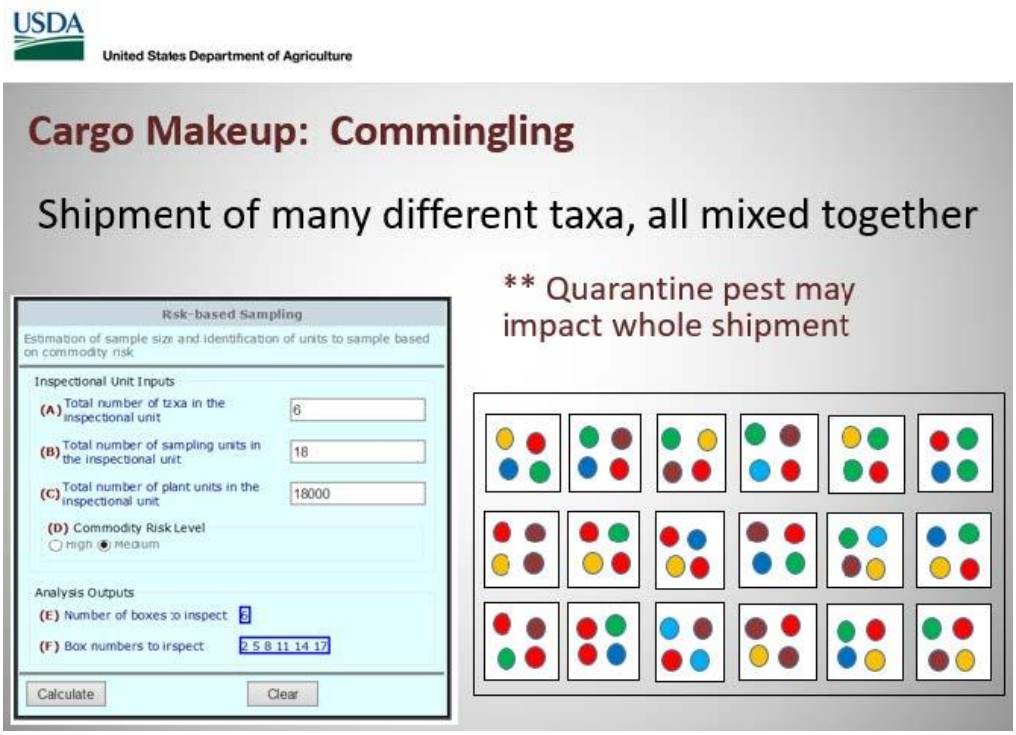


Figure 3. *Comingled shipment; each colored dot represents a different taxon in the shipment.* Source: https://nappo.org/application/files/5415/8676/4129/RBS_Symposium_Proceedings_-10062018-e.pdf

Imagine for example a shipment of flower bouquets with mixed roses and carnations. Since different products in the same sample unit cannot be separated, the challenge with comingled consignments is deciding on the sample rate. It is logical that the sample rate for the highest risk product would be selected, but this is likely to result in higher than usual interception rates on the lower risk products as a result of subjecting them to more rigorous inspection. These differences need to be taken into consideration when analyzing results and considering any changes in risk designations for the products in question.

2.9 Interagency cooperation

Phytosanitary authorities are not the only border agencies concerned with managing risk. Pest risk is only one aspect of the broader concern for compliance with national import requirements and can often be a lower priority even for agricultural imports.

Under the WTO Trade Facilitation Agreement (WTO-TF), which came into force in 2017, all WTO member countries **have agreed** that their national Customs Service will be the lead agency coordinating border operations and implementing risk-based policies. As countries adjust their national import designs to these relatively new obligations, there are multiple opportunities for NPPOs to redesign their border operations to gain future efficiencies. Two key areas to consider for RBS are data collection and randomizing cargo.

In the case of data collection, the advantage provided by the WTO-TF comes from the **single window** concept and the move toward digital data. The single window brings all relevant information on a consignment together in a single entry, a single system, and in electronic format. The single window greatly simplifies the entry process for all trade and provides a much more efficient way for border agencies to manage the trade clearance process. The key for NPPOs will be to proactively work with their Customs service to design electronic systems that collect, store, and make available relevant information in the single window, including feedback mechanisms for identifying inspection specifications, regulatory actions, and consignment status.

Establishing devanning routines that are coordinated with NPPOs and other border agencies provides benefits to multiple agencies and reduces negative impacts on trade.

The point about randomizing cargo is important because it is often highly impractical to unload containerized cargo for a full random inspection. However, if devanning (=unloading) is coordinated with the Customs service and other border authorities, the opportunity for random phytosanitary inspections can be increased. It is not likely that an inspection done for drugs or other prohibited articles will be

limited to samples taken from the rear of the container. Establishing devanning routines that are coordinated with NPPOs and other border agencies provides benefits to multiple agencies and reduces negative impacts on trade.

2.10 Stakeholder communication

The stakeholder base for border inspections is broad and diverse, but typically very focused on one thing: cargo clearance. Timely border clearance makes for efficient business processes and results in cost savings. Where RBS is concerned, it is crucial that stakeholders are informed about the advantages associated with its implementation, especially in the way that it is fair and predictable to trade. Consistently low risk consignments are cleared quickly and thus reward stakeholders for consistent regulatory compliance. Another point to highlight is that RBS designs are more difficult "to game" and therefore the distinctions between "dirty" and "clean" consignments will be more visible and actions more defendable from a risk standpoint.

Inspection of grape bunches from Perú by inspectors from the NPPO of Costa Rica.

3. DATA AND TOOLS

Robert Griffin[1] Maribel Hurtado[2] and Steve Hong[3]

1. National Coordinator for Agriculture Quarantine Inspection USDA, APHIS, PPQ – Retired
2. Scientific Officer NAPPO
3. Quantitative Analyst USDA, APHIS, PPQ

The type, amount, and quality of data collected by NPPOs, and the mechanisms for storing, sharing, and analyzing data for regulatory decisions vary greatly from country to country. The diversity of national situations creates challenges for operational harmonization but should not be a barrier to realizing the common objectives of free, fair, safe, and fast trade within the outlines of the international regulatory framework of the WTO and IPPC. The importance of inspection in this context cannot be overstated but the ability to effectively manage inspection requires the ability to measure it. These measurements require good data and appropriate tools to analyze, visualize, and compare key trade parameters, especially pest risk. International standards for phytosanitary measures (ISPMs) provide a starting point for addressing this need. The following discussions offer additional guidance to supplement the ISPMs and support operational implementation.

3.1. Data

Data provides the foundation for the analyses that support RBS. Better data offers better opportunities for meaningful analyses, but data also requires resources and effort to collect, store, and share. Resources and effort in a trade environment translate into costs. For this reason, the objective should never be to only collect data but rather to collect key data that are practical and most useful for analysis. Beyond this, there are important questions of data quality which include accuracy, reliability, and timeliness. All these aspects of data argue for thoughtful approaches to data collection for RBS.

Because every NPPO will have different situations and priorities that define their unique RBS interests, it is not possible to provide a generic outline for data collection. What this section aims to do is help NPPOs identify the most common data elements, understand their significance, and consider their relationship to analyses for RBS.

3.1.1. Types of data

In general terms, there are three categories of data:

1. Data regarding the consignment,
2. Data regarding the phytosanitary status of the consignment, and

3. Data regarding relevant regulations, policies, or administrative procedures.

The first data category provides the information needed to describe/distinguish a consignment; its products and quantity, its origin and destination, its value and other data required to satisfy border entry and business requirements. This may include a Phytosanitary Certificate (PC) if necessary, an invoice, a customs entry declaration, and various other documentation, some of which may be redundant. One of the main objectives of the WTO-TF is to eliminate redundancy and reduce this documentation burden to its minimum in a single window system that is administered by the national customs service. For NPPOs, this means that collaboration with customs will be crucial to establishing documentation requirements and gaining access to relevant information. As the single window becomes digital, it will be increasingly important for NPPOs to work closely with their national customs service to design electronic processes that collect, store, and share essential import data in a timely way with NPPOs.

The second data category on the phytosanitary status of the consignment is determined by the NPPO based on their own regulations, policies, and operations. For instance, the phytosanitary status of a consignment of frozen peas may depend only on verifying that the consignment is in fact frozen peas! On the other hand, a consignment of fresh cut flowers is likely to require inspection and may be rejected, treated, or subjected to other measures depending on the phytosanitary requirements of the (NPPO of the) importing country and on the results of inspection. In the case where a regulated pest is found, there will additional data collected by the NPPO to identify the pest and indicate what regulatory action was taken.

The final category of data refers to the regulatory and administrative situation surrounding each consignment. For instance, the importation of a consignment of treated fruit will have a different (lower) risk position in RBS than the same fruit imported without treatment. In essence, they become two different commodities for purposes of assigning risk.

Most of the data needed for RBS is transactional in nature, which means that it is related to a business activity, which is, the commercial import of regulated goods from one country to another. This raises two important points about data. The first is that there will be business information involved which can have legal and financial implications. Fortunately, this information is usually not important for RBS, but NPPOs need to be able to recognize such information and handle it as required. The second point is that import transactions typically involve multiple border agencies who have their own requirements and corresponding data needs, especially the national customs service.

In most cases, the information required by a national customs service, combined with information contained in a Phytosanitary Certificate (PC) will provide the data elements needed by NPPOs regarding the consignment. On rare occasions, NPPOs may need a data element that

is not available in this documentation. For instance, common names rather than scientific names may be used to identify products/commodities. Each NPPO will navigate their own data requirements in coordination with customs officials and other relevant border agencies to ensure the availability of essential data but also to prevent duplication of efforts and ensure that no unnecessary data requirements are created.

3.1.2. The nature of risk

Risk-Based Sampling aims to use the results of inspection to improve inspection by identifying the magnitude of risks across imported consignments so that adjustments can be made in inspection resources to maximize the effectiveness of risk management within the available resources.

Regulatory actions taken against consignments because of inspection findings are used as a proxy for risk. It is important to recognize the underlying assumption here; every regulatory action may be an equivalent observation from a data standpoint, but all pests are not equally risky.

Risk-Based Sampling aims to use the results of inspection to improve inspection by identifying the magnitude of risks across imported consignments so that adjustments can be made in inspection resources to maximize the effectiveness of risk management within the available resources.

A Pest Risk Analysis (PRA) is required to understand the true risk associated with each regulatory action. Since it is not practical to a perform PRA for every import-pest scenario, we assume that each regulatory action that results in the application of a phytosanitary measure against a consignment represents an average risk. This greatly facilitates analysis, but it also creates possible hazards. Imagine that two different commodities have an equal number of regulatory actions for the same number of consignments, but one is for a low- risk pest and the other is for a pest that is much more dangerous. RBS data would suggest that the two commodities are equivalent in their risk when, in fact, they are quite different because the pests represent vastly different risks.

3.1.3. Identifying data needs

The analysis of RBS data compares the rate of regulatory actions (as a proxy for risk) with some inspection parameter for which we have comparable data over a fixed period. The most common situation is to correlate regulatory actions with the commodity and its origin. A hypothetical example of the above would be, the number of regulatory actions on consignments of blueberries from Guatemala during the calendar year 2020.

The data requirements for this example are quite simple; we would need to know the number of regulatory actions taken for pests detected in blueberry consignments during the year 2020 and the number of consignments of blueberries from Guatemala that were imported into our country. This example demonstrates the simplicity of the relationship we are using as a measurement; however, the reality of our interest extends far beyond a single commodity (in this case, blueberries). To protect the plant resources of our country, we usually want to track and compare the actions taken on multiple commodities over multiple years or perhaps during specific months within a year. We also may be interested in comparing one country with another (that are sending us the same commodity) or understanding which suppliers within a country are causing the most problems (e.g., shipping the most infested consignments). We might also be interested in comparing the risk posed by different shipping pathways (e.g., air versus maritime) or we might like to examine the consistency of inspection results for the same commodity (ies) at different ports of entry. In sum, a wide range of analytical options are possible if the data is available, and the data is consistent (based on RBS).

3.1.4. Essential and non-essential data

As stated earlier, most of the data needed for RBS can be extracted from routine documents associated with inspection activities. Some important data elements are discussed below:

- **Consignment number or ID:** This is a unique identifier that links the consignment with the data record. It is essential in all cases as the means to establish an independent record.
- **Certificate number:** Phytosanitary Certificate (PC) numbers are important for notification purposes but are not essential for RBS unless the NPPO is interested in correlating risk to PC versus non-PC consignments, in which case the PC number is not as important as the number of PCs.
- **Date (dd/mm/yyyy):** The date of entry or inspection is an essential data element needed to identify consignments within defined periods of time and identify changes over time or pinpoint the seasonality of risk.
- **Import/Export:** Whether the inspection is for import or export is only important if the NPPO maintains data for both.
- **Lot size N:** The lot size is necessary to calculate the sample size and may need to be recorded for official purposes depending on the regulatory policies of the NPPO.
- **Sampling unit:** The sample unit (box, bag, piece, etc.) is important to record only once if it is consistent or in every instance if it varies within the consignment. The key point is to ensure that equal data can compared.
- **Sample size:** The sample size is essential data if it varies from the predetermined detection level for sampling. For example, if all sampling is done for a 5% detection level, then the sampling will be consistent, but any changes made to the sampling (such as an

"extra box") changes the statistical significance of the results and should be noted as a deviation.

- **Randomization of samples:** In a best-case scenario, sampling will be completely randomized but this is rarely possible. When it is possible, it is essential to record it because it represents an important data point for comparison. In the case of less-than random sampling, there should be either a general record or record for each consignment that describes the level of randomization. This is important for understanding the level of confidence associated with inspection results.
- **Country of origin:** Essential data for imports.
- **Country of destination:** Essential data for exports.
- **Product (common name):** Common names are conventional on import documents but often are not useful and can be misleading for plant health (NPPO) purposes. Common names are essential when scientific names are not provided because they assist in properly identifying the commodity.
- **Pathway:** Identifying whether the consignment is arriving by air, sea, or land and whether the commodity is for consumption or propagation is considered essential data because it facilitates important risk comparisons.
- **Product (scientific name):** Scientific names and even varieties can be especially important for recording the pest-host relationship. This is essential data for pest records.
- *Product category*: Indicating whether the commodity is fresh or processed, a fruit, vegetable, seed, plant for planting, flower, etc. is essential for risk comparisons.
- **Importer:** This is essential data for linking the consignment to a destination.
- **Exporter:** This is not essential data but can be useful for comparing sources.
- **Producer:** The producer is a key risk factor and can be essential data when distinguishing individual high-risk sources from among many commodity suppliers.
- **Pest common name:** The common name of pests is non-essential and may lead to confusion if different countries use different common names or use the same common name for different pest species.
- **Pest scientific name:** Essential data.
- **Pest type:** The type of pest that was intercepted (e.g., insect, mite, nematode, weed, mollusk, etc.) is essential.
- **Pest stage:** The life stage of the pest (e.g., egg, nymph, larva, pupa, adult) is essential data.
- **Number of pests:** The total number of a specific pest found in the consignment is essential for understanding the infestation rate. Recording no (or zero) detections is also important (see section 3.1.5 below).

- **Pest categorization:** Data regarding whether the pest a quarantine pest (regulated), a non-quarantine pest (non-regulated) or a regulated non-quarantine pest is non-essential data except to note that only regulated pests are actionable.
- **Pest risk level:** A record of whether the pest risk is high, medium, low, or negligible is important only if the NPPO maintains/uses this classification.
- **Action:** The phytosanitary status of the consignment is the ultimate determining factor for its place as an inspection record. Consignments requiring no regulatory action are equally important inspection records (see section 3.1.5).
- **Name of inspector performing the inspection:** This is non-essential data unless the NPPO is tracking this information for other purposes than RBS (e.g., inspector performance ratings).
- **General observations:** Non-essential data except as determined by the NPPO.

3.1.5. Collecting zeros

Recording an inspection with an observation of zero is crucial for the implementation of RBS. A result of zero for "number of pests found in the consignment" means that no pest was intercepted during the inspection process. Likewise, a zero for the variable regulatory action indicates that the consignment was not subjected to a risk mitigation measure. The number of pests and regulatory actions become meaningful by understanding their relationship to the total number of inspections yielding zero's. For example, compare one action from 10 inspections versus one action from 100 inspections. The difference between a 1% action rate and a 10% action rate is only visible by knowing the number of inspections that had no action. Likewise, if inspection data for a specific commodity from a specific country of origin shows zero actions for a defined period or number of consignments, one can reasonably conclude that importing this commodity from this origin presents a very low or negligible pest risk.

3.1.6. Data shortcomings

Data is not useful for analysis if it is insufficient or lacks the necessary quality. There are two important strategies to consider here. The first strategy is ensuring that documentation associated with consignments provides appropriate information. Most of this data goes to Customs and is associated with the official import entry. This transactional data is normally very consistent but can also be inadequate. For instance, documentation accompanying a consignment of cut flowers may not identify the different species – information that could be important to the NPPO from a risk standpoint.

The second strategy involves the phytosanitary data that results from inspection. This is where the bulk of responsibility falls on the NPPO to understand its risk management objectives and priorities so that it can design appropriate mechanisms to collect, store, distribute, and analyze

data on the phytosanitary status of consignments. The ability of the NPPO to effectively manage pest risk will depend primarily on the data it collects and the quality and quantity of this data, so it deserves careful attention, especially when beginning to use or when transitioning to RBS.

Historical data from the transactional category (number and size of consignments, origin, type of commodity, etc.) is often easily accessed and it can be tempting to believe there are opportunities to use it for RBS. A critical question that must be answered before attempting such analyses is whether the corresponding/associated phytosanitary data is also appropriate. If the historical inspection data is not based on sampling for a consistent level of detection, the results are only weakly correlated to the transactional data for purposes of evaluating risk. Some broad conclusions may be drawn from relative relationships of regulatory actions to consignments, but the data lack analytical validity for adjusting import program parameters or requirements.

Generally, NPPOs initiating their RBS programs using historical data should begin with data sets that have the greatest consistency and focus analyses on a few elements (commodity, pest, origin, etc.) to gain experience while also collecting more and better data for expanded analysis in the future.

3.2. Tools

The central question for NPPOs to ask about their inspection data is, what does it show? The converse question is just as important, and that is - what cannot be shown with the inspection data we have? An array of different tools and analytical techniques are available to facilitate data manipulation and analysis, but care is required to avoid the temptation to immediately choose the most sophisticated solutions before establishing a firm understanding of the limits of the data, tools, and analyses. What is perhaps most important is understanding the limitations of experience and expertise with some of the more sophisticated approaches. The aim in this section is to describe the basic analytical tools and provide guidance on their use.

3.2.1. Microsoft Excel

Data has no value for analysis unless it is in a format that allows it to be easily stored and accessed. Microsoft Excel is a useful tool for the collection and organization of data into one or more workbooks. Excel is commonly linked to larger data collection sources and can be used to store, access, filter and search inspection data. NAPPO has developed specific workbooks in Excel that can be adapted by each NPPO to assist in the collection and organization of inspection data. The workbook known as the database for inspection data, includes most of the categories and fields that are useful for RBS and that were described earlier in this chapter. The workbooks can be downloaded at no cost from the NAPPO website.

In addition to accepting numeric as well as text/written data, the cells in Excel can also contain embedded formulae that facilitate basic calculations as well as increasingly complex mathematic, trigonometric, arithmetic, financial, logic or statistical functions. Excel can be used for basic data analysis and to generate graphs based on the data contained in its workbooks.

Additional Excel functionalities such as pivot tables and macros facilitate analysis of large amounts of data without needing to develop the formulae "de novo". These functionalities can facilitate automated data analysis.

Some of the analyses that can be performed with inspection data using Excel include, among others, analysis of variance, correlation calculations between two variables, covariances, generation of descriptive statistics reports, frequency histograms, and generation of random numbers. On the other hand, Excel allows data organization in order to carry out these and other analyses in different statistical programs such as R.

3.2.2. Hypergeometric tables

Hypergeometric tables are a useful tool to determine the sample size to be inspected from a given lot or to determine the acceptable risk level for an already inspected sample. They provide the hypergeometric probability distribution for the sample size in different lot sizes with different acceptable risk and confidence levels. Likewise, they make it possible to determine the acceptable risk level given the parameters of lot size, confidence level, and sample size. The hypergeometric tables for different lot sizes (100 to 200,000) can be found and freely downloaded from the NAPPO website.

3.2.3. Sample size calculator

It is also possible to calculate the appropriate sample size for inspection without using hypergeometric tables. A sample size calculator created in Excel and freely available from NAPPO (see below) uses the formula below developed by Fosgate for surveillance and animal disease detection and to document disease freedom after an outbreak (Fosgate, 2009). The use of this formula for RBS purposes requires determining the pest prevalence that the NPPO considers important to detect (= acceptable level of risk), defining the desired level of confidence, and knowing the lot size. The formula to calculate sample size is as follows:

$$\mathbf{n}=[1-(\alpha)^{1/d}]\ \{N-[(d-1)/2]\}$$

Where α is 1 – the level of confidence, N is the population size or the lot size, and *d* is the expected number of pests in the population.

The sample size calculator and the step-by-step instructions on how to use it, are available on the NAPPO website.

1	***Detection level Risk acceptance level***	**10%**
2	***Confidence level (1-a)***	**0.95**
3	***Lot size (N)***	**100**
4	***Sampling unit***	**Box**
	Sample size (n)	**25**

Figure 4. *Sample size calculator.*

An example is provided above (see **Figure 4**) using a detection level of 10% [1], a confidence level of 95% (0.95) [2], a lot size of 100 [3], and the sampling unit which will be boxes [4]. Using these inputs, the calculator will indicate that the sample size to be taken corresponds to 25 boxes. In this example a lot of 100 boxes would require 25 boxes to be inspected to achieve a detection level of 10% with 95% confidence.

3.2.4. Detection level calculator

A detection level calculator is being developed by NAPPO. This calculator is intended to be used to determine the detection level or acceptable risk level for an inspection when the confidence level, lot size, sampling unit data, and sample size are provided. The ability to calculate the detection level from sampling information is useful for understanding the level of detection for non-RBS inspections and the range of detection levels for different inspections. This information supports the analysis of existing inspection designs and identify areas of greatest concern for risk-based adjustments.

3.2.5. R and R Studio

R is a freely downloadable software environment designed for statistical computing and graphics. R includes an effective data handling and storage capability; a suite of operators for calculations; an integrated collection of tools for data analysis; visual tools for data analysis and display and uses simple and effective programming language. R is available under the terms of the Free Software Foundation's GNU General Public License in source code from: https://www.r-project.org/

R Studio is freely downloadable here https://www.rstudio.com/. It is an integrated development environment (IDE) for R.

In terms of RBS, R and R Studio can be used to perform more in-depth analysis and interpretation of properly collected inspection data. In addition, R and R Studio can also be used to visualize inspection data through the generation of *Treemaps* which are explained below.

3.2.6. Treeemaps in R

A *Treemap* is a visualization tool used to display hierarchical data using nested rectangles in a tree-like structure. It greatly simplifies the understanding of the relationship between data variables, including inspection data. A *Treemap* captures two types of information: the value of individual data points and the structure of the hierarchy generated by the data. The area of each generated rectangle is proportional to its value. *Treemaps* are generally used when wanting/needing to visualize proportions derived from a large amount of hierarchical data.

The *Treemap* function in R allows the user great flexibility when drawing a *Treemap*. In the example below (**Figure 5**), hypothetical inspection data on action rates for different consignments (flowers, walnuts, mango, coffee, Palm oil, etc.) from different countries (Countries 1 – 10) were used to generate a *Treemap* using R. The color intensity in the *Treemap* rectangles indicate higher regulatory action rates for the commodities. The size of each commodity rectangle represents the volume of that specific commodity imported from that specific Country (Mango from Country 1 versus Mango from Country 10) into your Country.

The R *Treemap* visualization tool allows for an easy assessment of where pest risk is higher in this hypothetical example. For example, it is evident that most commodities exported from Country 7 have the highest regulatory action rates following inspection, followed by those originating in Country 1. It is also clear that.

- Not all Countries that export mango (to your Country) present the same level of risk even though they export different volumes of mango.
- Palm oil has very low regulatory action rates irrespective of Country of origin.

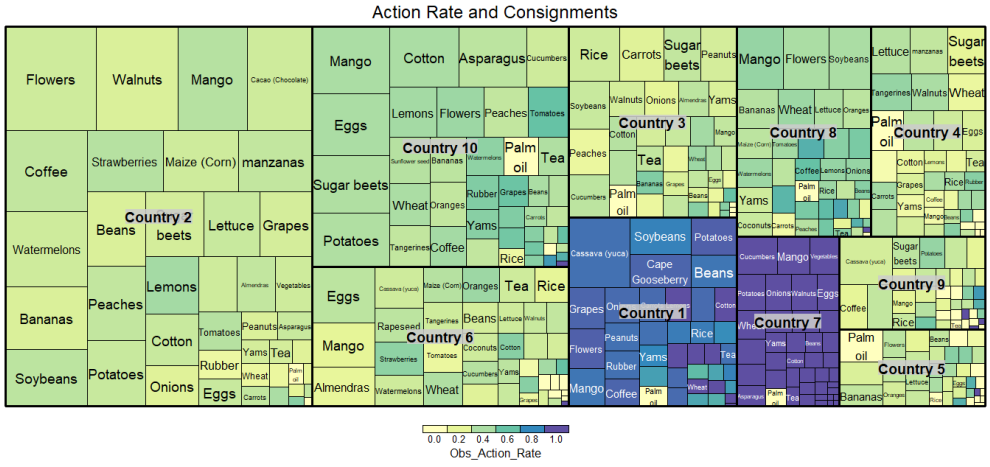


Figure 5. *Treemap representing regulatory action rates for different consignments, see above for explanation.*

The step-by-step instructions on how to develop *Treemaps* using R and properly collected inspection data will soon be available to be freely downloaded from the NAPPO web site.

3.2.7. Using RBS for data analysis and risk categorization

Recently, Kim *et al.* (2018) used inspection data from the U.S. Department of Agriculture (USDA) to estimate the probability of the presence of quarantine pests on propagative plant materials imported from different countries and develop a risk-ranking methodology for the different country–commodity combinations.

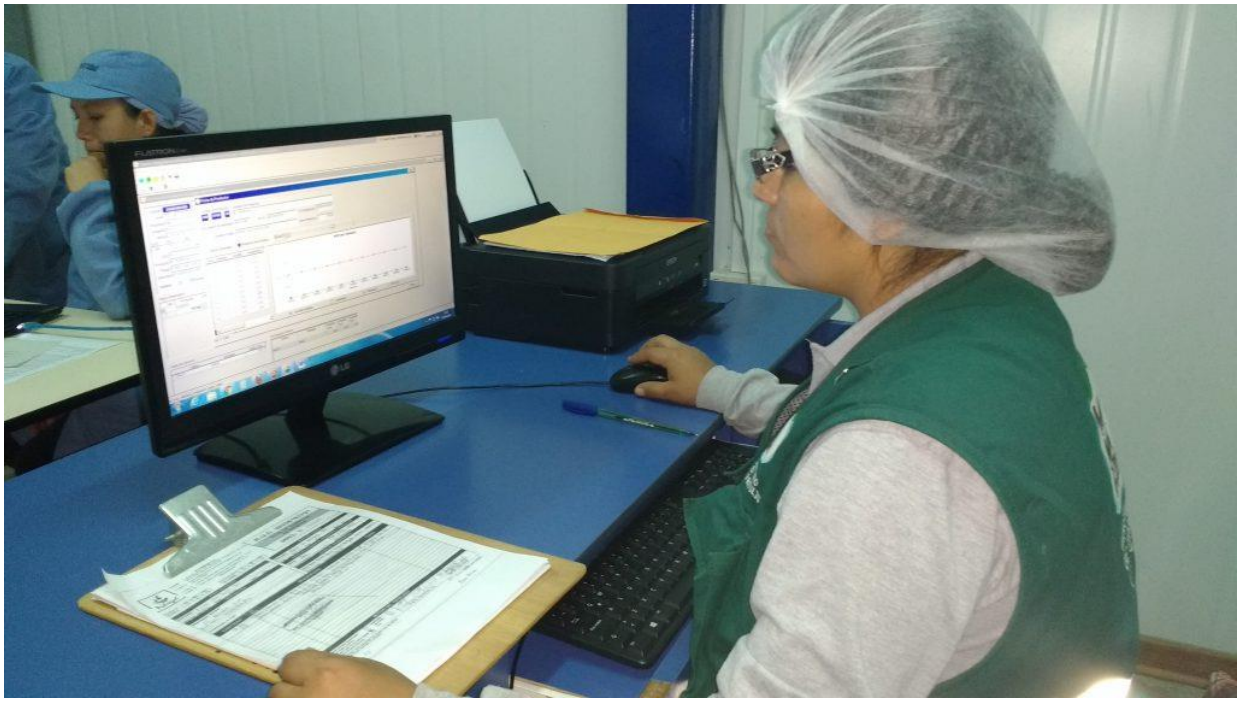
They chose to analyze inspection records for propagative material because importation of plants for planting poses a higher risk than imports of other regulated plant products and because a history of pest interception data was available for analysis from the USDA. The inspection data was split into two data sets (70% - training data set and 30% - test data set) which were used to develop the predictive models (using the training data set) and conduct the validation study (using the test data set).

Kim et al (2018) used a generalized linear model (GLM) with Bayesian inference and a generalized linear mixed-effects model (GLMM) contained in R version 3.2.5 to estimate interception rates on different country-commodity combinations and their associated uncertainties.

They categorized country-commodity combinations into different compliance levels based on simulated interception rates of quarantine pests and predetermined thresholds and compared the categorization results among models. They used a two-step approach in their analysis. First, country–commodity combinations were separated into small and large variance groups based on confidence intervals of the estimated probabilities of carrying quarantine pests. Second, each group was further partitioned into different compliance levels (High, Medium, Low, and poor/unacceptable) using defined thresholds.

Analysis of the data determined the top five plant genera carrying quarantine pests, revealed the most frequently imported plant genus, and highlighted which country-commodity combinations had the highest number of quarantine pest interceptions.

The study found that the prediction ability of the GLMM was greater than that for GLM. However, predicted interception rates and their confidence intervals were influenced by the statistical models used. This suggests that care must be taken when applying results such as these to the development of inspection programs with different monitoring intensities based on compliance levels of country-commodity combinations.



Entering inspection data.

Source - https://www.senasa.gob.pe/senasacontigo/ica-inspeccion-fitosanitaria-de-palta-hass-para-exportacion-a-china/

4. SAMPLING METHODS

Andrew Robinson[1], Mark Ducey[2], Raphael Trouve[3], Marta Elva Ramírez Guzmán[4], Robert Griffin[5], Steve Hong[6]

1. Director, CEBRA, School of Biosciences - Reader & Associate Professor in Applied Statistics, Australia
2. Professor - University of New Hampshire USA
3. Post-Doctoral Research Fellow - CEBRA Australia
4. Professor of Statistics - Colegio de Postgraduados Mexico
5. National Coordinator for Agriculture Quarantine Inspection USDA, APHIS, PPQ - Retired
6. Quantitative Analyst USDA, APHIS, PPQ

The goal in sampling applications is to be able to make a statistical statement about some characteristic of a collection of objects. For example, we may wish to estimate the average family income in a city, or how many families in the city have an average income that is below some threshold. We may wish to know the number of habitat trees in a forest, or the total woody volume in a forest. We may wish to estimate the proportion of contaminated fruit in a consignment of oranges that has arrived at our national border, or the proportion of orange trees in an orchard that are infested with a pest. In all these instances, it is too expensive and too slow to inspect or measure all the population, so we resort to taking a sample from the population, measuring the units in the sample, and drawing a conclusion about the population from the sample.

In order to draw a conclusion about the unknown population that is based on the known sample of units, it is essential that the sample be taken correctly, and that the data arising from the sample be analyzed correctly. Sampling theory guides us in how to do both things. Failure in either step will yield conclusions that do not have the intended statistical credibility and will therefore potentially mislead the analyst or the decision-maker. Sampling theory allows us to learn about the whole population without measuring the whole population. Correct and efficient use of sampling theory requires an appreciation of the relationship between the population, the sampling units that comprise the population, and how to select the sample.

Sampling theory allows us to learn about the whole population without measuring the whole population.

4.1. Sampling unit, population, and frame

Sampling units are the units into which the population is divided, and which are selected and then measured in some way. The definition of the sampling units and the population are

therefore linked (see **Table 1** below). In an operational setting, the choice of sampling unit (e.g., container, box, item) will depend on the physical as well as the biosecurity context, including the size of the items comprising the consignment (e.g., watermelons, apples, or strawberries) and the nature of the pest (e.g., mobile insect, sessile insect, fungal pathogen).

In order to select a sample of units from the population we require a frame, which in its simplest form, is a list of numbers corresponding to each unit in the population. Sampling works by randomly selecting the units from the frame and then measuring the sampling units in the population that correspond to the units in the frame. Sometimes the frame is an imperfect match to the population, and unmeasurable but hopefully small errors ensue. Example frames, some imperfect, are also included in **Table 1**.

Table 1. *Example of general and plant health scenarios, populations sampling units and frames.*

Scenario	Population	Sampling Unit	Example Frame
Estimate the average family income in a state.	All the families in the state.	A family.	State taxation records.
Estimate the number of Total trees in a forest.	The land area covered by the forest.	A square 0.04 ha plot.	A map of the forest area.
Estimate the proportion of contaminated fruit in a consignment of oranges.	All of the orange crates/boxes in the consignment.	A crate/box of oranges.	A list of the boxes or a list of numbers from 1 to the box count.
Estimate the proportion of contaminated fruit in a consignment of oranges.	All of the oranges in the consignment.	An orange.	A list of numbers from 1 to the inspected orange count.
Estimate the proportion of orange trees in an orchard that are infested with a pathogen	All of the orange trees in the orchard.	An orange tree.	A list of numbers from 1 to the infested tree count.

In the border inspection setting, sampling is typically applied to consignments of regulated articles, and inspection is performed to determine whether the consignments are compliant with biosecurity regulations. If the sample detects a non-compliance, then further regulatory action is taken, for example treatment, re-export, or destruction, whereas if no non-compliance is detected then the consignment is typically released into commerce. Consequently, sampling and inspection for biosecurity compliance are different than the traditional sampling applications, because the objective is to make a decision (release/intervene) instead of only making an estimate (e.g., of the contamination rate).

Regardless of sampling objective, we should specify the sample size – the number of units in the sample – before we begin sampling. The approach to determining the sample size applies to estimation, rather than decision-making. We try to choose a sample size that will provide an estimate that isn't too uncertain, recognizing that increasing the sample size decreases the uncertainty but also increases the time and cost associated with sampling.

Sampling and inspection for biosecurity compliance are different than the traditional sampling applications because the objective is to make a decision (release/intervene) instead of making an estimate.

To choose a sample size for biosecurity compliance, we assume that the only condition under which we will release the consignment is if an inspection of *n* units results in no pest detections/interceptions, let's say x = 0. Then, we choose *n* in such a way that we are unlikely to detect no pests in *n* inspections unless the proportion of infested units is extremely low. Then we can compute the sample size. For example, if we want to detect pests with 95% probability in a consignment for which 0.5% of the units are infested, then we must inspect about 600 units, assuming that inspection will always detect pests that are present. See additional information in Sections 4.2.1. and 4.4.3.

There is an additional consideration, which is that sampling may also be applied within pathways, that is, the consignments that are inspected may themselves be selected from a pathway of consignments. So, we may have a random sample of consignments taken from the pathway, and within the chosen consignments, we may select a random sample of units, for example oranges, or crates of oranges. We would do this if we were interested in monitoring the pathway in case the biosecurity risk might change.

In each case, having decided upon the population and sampling unit, and having obtained the best frame available, we then choose the sample. There are many ways of choosing samples (sample designs) that rely on the availability of different types of information, or different expectations about what the population will be like. We will cover some of these in the next section. In the absence of any other information, the most informative sample is one that selects the units completely at random – the Simple Random Sample.

4.2. Sampling methods

Below is a brief overview of different kinds of sampling methods. The methods are split into two classes: statistically based sampling – where sampling outcomes have known statistical properties – and non-statistically based sampling, where is sampling done in some other way.

4.2.1. Definitions and related concepts

In statistically based sampling, we develop the sample design, collect the sample, and analyze the data according to a prescribed recipe. Below are definitions important to statistically based sampling.

Parameter – the population parameter is the characteristic that we wish to estimate. For phytosanitary inspections we design the sample as though we want to estimate the infestation rate in the population.

Sample size – the sample size is the number of units selected from the lot or consignment that will be inspected or tested (FAO, 2016). This is usually denoted by the letter n, as indicated earlier.

Level of detection –the level of detection is directly related to the confidence level; in that it is the lower limit of the proportion of infested units that should be detected at the given confidence level. We will use *p* for pest prevalence.

Confidence level – the confidence level is the minimum probability at which we wish to detect infestation in the sample, given that the baseline infestation rate is equal to or higher than the level of detection. In the example above (see 4.1), the confidence level was 95%. The confidence level is also the *sensitivity* (= *s*) with which a test detects a negative outcome.

ISPM 31 says: "95% confidence level means that the conclusions drawn from sampling will detect a non-compliant consignment, on average, 95 times out of 100, and therefore, it may be assumed that, on average, 5% of non-compliant consignments will not be detected." (IPPC 2008). This statement is true, however, in the present context, it may be misleading because of the definition of non-compliance. In the example above (see 4.1) the probability of detecting a large consignment with exactly 0.5% infestation was 95% when sampling (about) 600 units. If the infestation were higher – and therefore still non-compliant – then the probability of detection would be higher than 95%. For example, the probability of detecting a 1% non-compliant consignment using a 600-unit sample would be about 99.8%. As such, it is important to indicate the level of non-compliance along with the detection confidence level. See additional information in Section 4.4.3.

Efficacy of detection – it is possible that an infested unit will be inspected but the infestation may not be detected. If we allow for this possibility, we do so by using the efficacy of detection, which is defined as the probability of detecting infestation that is present. In the example above (see 4.1) we assumed "perfect" detection, so the probability and therefore the efficacy of detection were equal to 1.

4.2.2. Statistically based sampling methods

This section describes commonly used statistically based sampling methods.

a. **Simple random sampling (SRS or SyRS)** is the selection of a sample of size *n* from a process such that every possible combination of *n* sampling units has the same probability of being selected. This definition is more stringent than simply stating that each *n* unit has the same probability of being selected. The SRS design is the foundation of sampling theory but is rarely used. An example of SRS in biosecurity is the automated computer selection of consignments for inspection with a set probability.

b. **Systematic sampling (SyS)** is an alternative to SRS that imposes a grid on the process, selecting every *k*-th unit. Two things are needed to employ SyS: a grid spacing *k*, computed to achieve a desired sample size *n*, and a random starting point. SyS has an advantage for inspection of consignments where pests are clustered, as the probability of detecting one pest cluster is higher for SyS. Wolter (1985) has a comprehensive review of SyS.

c. **Cluster sampling (CS)** is applied when sampling is hierarchical, and it is easier and less costly to sample sets of units than sampling the units themselves. For example, for apples packed in boxes, it may be more convenient to randomly or systematically select the boxes and inspect all the apples than to randomly or systematically select the apples within the boxes. In this example the boxes are treated as though they were clusters of apples. CS achieves nominal sensitivity for randomly dispersed pests, but not for clustered pests, for which the sensitivity will be lower than desired.

d. **Stratified sampling (StS)** is a way of organizing the sampling units before they are selected. In StS, all the units are classified into strata and then each stratum is sampled as though it were unique. StS requires additional information about each unit, is used to allocate it into the appropriate stratum.

e. **Sequential sampling (SeS)** involves a change to the stop rule, which indicates how many units the sample should have. Earlier we indicated that sample size is set before sampling begins. Here, the stop rule is that If n samples have been taken, then sampling is stopped. This is known as fixed *n* sampling. Under SeS the stop rule is different. Sampling typically continues until a *statistical condition* is met. SeS has lower strategic value than fixed n sampling because the estimator of prevalence is less efficient. However, if the sole purpose is to determine the biosecurity status of items in the pathway, then SeS may be preferred. Note that if the sample units are correlated, instead of being independent, then SeS can perform poorly (Robinson and Hamann, 2008).

f. **Fixed proportion sampling (FPS)** involves sampling a specific proportion of the consignment, for example, 2%. FPS yields inconsistent levels of detection or confidence as consignment sizes vary. That is, when the set level of detection is expressed as a percentage of the consignment, FPS achieves it with confidence that depends on consignment size. However, if the set level of detection is expressed as an absolute number, then FPS confidence is independent of the consignment size.

4.2.3. Non-statistically based sampling methods

This section describes commonly used non statistically based sampling methods.

a. **Convenience sampling** according ISPM 31, involves selecting the most convenient (for example, most accessible, less costly, fastest) units from the consignment, without selecting units in a random or systematic manner (FAO , 2008).

b. **Haphazard sampling** involves selecting arbitrary units without using true randomization. Haphazard sampling may appear to be random because the inspector is not aware of having a sampling bias. However, unconscious bias may occur, so that the degree to which the sample is representative of the lot is unknown (FAO , 2008).

c. **Selective or targeted sampling** according ISPM 31, involves deliberate sample selection from parts of the consignment most likely to be infested, or from units that are obviously infested, to increase the chance of detecting a specific pest. This method may be favored by inspectors who are familiar with the commodity and the pest's biology. Use of selected or targeted sampling may be triggered when wanting to identify a specific part of a consignment that has a higher probability of being infested. For example, a section of wet timber may be more likely to harbor nematodes that dry sections of timber. Because the sample is targeted, and hence statistically biased, a probabilistic statement about the infestation level in the lot cannot be made. However, if the sole purpose of sampling is to increase the chance of finding a regulated pest(s), this method is valid. However, additional sampling may be required to reach general confidence in the detection of other regulated pests. The use of selective or targeted sampling may limit the conclusions about the overall pest status of the consignment.

4.3. Implementing sampling plans

4.3.1. Sample design and selection

Sample design and selection should consider what knowledge is available on the distribution of pests in the lots or consignments. When pest distribution is unknown and because sampling is done without replacement and the population size is finite, hypergeometric distribution should

be used to determine the sample size. A hypergeometric distribution indicates the probability of detecting a certain number of infested units in a given sample size drawn from a lot of a given size, when a specific number of infested units exist in the lot or consignment (FAO , 2008). Hypergeometric tables are useful tools in this respect. Additional information on hypergeometric tables and their use can be found here.

When pest distribution in a lot or consignment is clustered or aggregated, ISPM 31 indicates that it will always lower the likelihood of finding an infestation (IPPC, 2008). However, Simple Random Sampling (SRS) described above in Section 4.2.2., achieves nominal sensitivity when applied to situations where pests are both randomly dispersed and clustered (Yamamura *et al.*, 2015; Lane *et al.,* 2019).

4.3.2. Multiple lots in a consignment

Often a consignment will be composed of several lots that are similar in some ways and different in others. For example, a consignment of oranges may include lots from more than one supplier. A question for this scenario might be: How can consignment-level assurance be achieved? If, for example, a 600-unit inspection is routinely applied to a single-source consignment (one supplier), then should a 600-unit inspection be used for each different supplier, or could we split the 600-unit sample in some way that would achieve consignment-level assurance?

ISPM 31 indicates that "Treating multiple commodities as a single lot for convenience may mean that statistical inferences cannot be drawn from the results of the sampling." (IPPC 2008). However, Lane *et al.* (2019) demonstrated that consignment-level assurance can be achieved by applying the usual sampling methods and can be improved by means of stratification with proportional allocation. This means that the sample needs to be allocated between the lots within the consignment proportionally to the number of units within each lot, with one important caveat - the entire consignment must be treated according to the outcome of the sample inspection.

To illustrate further, if a consignment includes lots A and B, with 20,000 and 40,000 units respectively, then nominal consignment-level assurance can be achieved if no infestation is detected from a sample of 200 and 400 units respectively. However, if infestation is detected in lot A only, then both lots must be subjected to the phytosanitary measures. This is because the stratification approach does not provide the same assurance than if each lot were treated as a unique consignment. In order to achieve that level of assurance, we would need to apply the 600-unit sample to each lot. However, when the goal is consignment-level assurance, the stratification approach is sufficient.

4.4. Probability distributions

4.4.1. Useful formulae

The probability mass function, sensitivity, and sample size formulae for the hypergeometric, binomial, Poisson, and beta-binomial sampling are given in **Table 2**.

Table 2: *Probability mass function, sensitivity, and sample size for different types of distribution.*

Type of sampling	Probability mass function	Sensitivity	Sample size
Hypergeometric	$Pr(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$S = 1 - \frac{\binom{N-K}{n}}{\binom{N}{n}}$	n ~ [1 – (1-S)$^{1/(pN)}$] [N – (pN-1)/2] $n \sim [1-(1-S)^{1/(pN)}][N-(pN-1)/2]$
Binomial	$Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	S = 1 – (1-p)^n	$n=\frac{ln(1-S)}{ln(1-p)}$
Poisson	Pr(X=k) = $\lambda^k e^{\lambda} / k!$ λ = pn	S = 1-e^{-pn}	$n=\frac{-ln(1-S)}{p}$
Beta-binomial	$Pr(X=k)=\binom{k}{n}\frac{B(\alpha+k,\beta+n-k)}{B(\alpha,\beta)}$ Where B is the beta function.	S ~ (1+nθ)$^{mp/θ}$ p = α/(α+β) ρ=1/(1+α+β) m = number of clusters = nk/n	$n \sim -\frac{n_k\theta}{p}\frac{ln(1-S)}{ln(1+n_k\theta)}$ n_k = number of samples per cluster θ= 1/(α+β)= rho/(1-rho)

Starting from the distribution's probability mass function Pr(X=k), *X* being a discrete random variable and *k* the number of infested samples, we can compute the sensitivity (confidence-level) of an inspection as the probability of detecting at least one infested unit in the inspection: S = Pr(X≥1) = 1 – Pr(X=0). We compute the sample size *n* required to detect a prevalence *p* with a confidence-level *S* by rearranging the sensitivity equation and fixing *p* and *S* to a given value (and, depending on the distribution, additional parameters such as lot size *N*, or aggregation index ϑ).

An alternative to controlling for the design prevalence and sensitivity of the inspection is controlling for slippage (leakage). Slippage is the number of infested units not detected during inspection and that end up entering the country. Slippage is calculated as the proportion of consignments that are deemed compliant (1 – S) times the number of infested units in these consignments p(N-n). Another useful metric is slippage rate defined as slippage divided by lot size: p(N-n)/N (see 4.4.5.).

4.4.2. Sample size for small lots: Hypergeometric-based sampling (simple random sampling)

When inspecting small lots, it is important to remember that sampling is done without replacement (e.g., we do not put an orange back in the lot after inspecting it). Sampling without replacement is described by the hypergeometric distribution. It has four parameters: the number of units *N* and infested units *K* in the lot, and the number of units *n* and infested units *k* in the sample (see **Table 2**). The lot infestation rate is p = K/N. While there is no close form solution for the sample size for the hypergeometric distribution, we can solve for *n* either by doing a simple search over the values of *n* or use the approximation given in **Table 2** rounded to the higher integer. When the number of units sampled *n* is relatively large compared to lot size *N* (say more than 5%), hypergeometric sampling instead of binomial sampling should be used as it allows reducing sample size without compromising the level-of-detection and the confidence-level.

4.4.3. Sampling of large lots: binomial or Poisson-based sampling

When the number of sampled units is much smaller than the number of units in the lot (say less than 5%), we can simplify calculations by approximating sampling without replacement (hypergeometric) with sampling with replacement (binomial). Binomial sampling is the most common sampling method in biosecurity and forms the basis for the 600-sample rule used in some biosecurity programs. The 600 sample rule detects 95% of consignments having an infestation rate of 0.5%, calculated as follows: S = 1 - (1 – 0.005)^597 = 0.95 and n = log(1-0.95) / (log(1-0.005)) = 597, and often rounded to 600.

Poisson-based sampling may be used as further approximation to binomial sampling when *p* is low and *n* is large. For example, Poisson sampling detects 95% of consignments having an infestation rate of 0.5% when *n* = -log(1-0.95)/0.005 = 599. Poisson sampling is seldom used in practice.

Arya, *et al.* (2012) and Fosgate, (2009) indicate that when the target population is very large (theoretically infinite), hypergeometric distribution can be well approximated by a binomial one.

4.4.4. Sampling for pests with an aggregated/clustered distribution: beta-binomial based sampling

When data is aggregated/clustered, we have to account for clustering in the statistical procedure. The most used model for clustered data is the beta-binomial model. The beta distribution can be parameterized in terms of parameters α and β, in terms of parameters *p* (mean infestation rate) and ρ (ICC), or in terms of *p* and θ (aggregation index) (see **Table 2**). It can be useful to switch between parametrizations when one type is more convenient for some the calculations.

Aggregation/clustering reduces inspection sensitivity (= reduces the chance of finding infested units during inspection). This means that if we want to keep the sensitivity constant, we need to increase the sample size relative to that of the binomial sampling. How much more to sample depends on the intra-cluster correlation coefficient (ICC or of the aggregation index θ) and the number of samples per cluster. When we sample one unit per cluster and the clusters themselves are sampled at random, the sample size is equivalent to binomial sampling. When the ICC is zero (= there is no clustering), we can also use the formula for binomial sampling. Approximate formulae for the sensitivity and sample size are given in Madden *et al.* (1996) and Venette (2002) and reproduced in **Table 2.** The approximation is based on the negative binomial approximation to the beta-binomial distribution when *p* is low, analogous to the Poisson limit to the binomial.

Computing the sensitivity and sample size for beta-binomial model (**Table 2**) requires an estimate of the ICC (or of θ). However, there is little guidance on how to obtain this estimate. A first approach is to fix the ICC to a reasonable value or to extract the value from the literature when estimates are available. For example, Madden *et al.* (1996) suggest θ of 0.016-0.090 for grape plants infected with the fungus *Eutypa lata,* and Hughes and Madden (1993) report θ of 0.0056-0.123 for tobacco virus in tobacco plants.

A more tailored approach to estimating the ICC is to fit a beta-binomial model to past inspection data from the pathway. A first estimate of the ICC can be obtained by using the method of moment to one inspection in the pathway $\theta = (s^2 - np(1-p)) / (n^2 p (1-p) - s^2)$, where s^2 is the sample variance or by maximum likelihood (Griffiths 1973). Modern statistical software such as R (see section 3.2.5. for more on R) provide packages to accomplish this (VGAM, GAMLSS). However, a single consignment may not have enough data and might not be representative of the entire pathway. A better approach is to fit a hierarchical beta-binomial model to all consignments in a pathway and accommodate for varying prevalence among different consignments. The equation for hierarchical beta-binomial model is:

$$k \sim betabinom(p_j, n, \rho)$$
$$logit(p_j) \sim N(\gamma, \sigma)$$

The top row represents a beta-binomial distribution of infested units among crates in consignment j with prevalence p_j. The bottom row represents the varying infestation rate among consignments. Fitting the ρ parameter on the logit scale allows it to be bounded in the 0-1 range.

To analyse how estimates of p and ρ in the hierarchical beta-binomial model are affected by the number of consignments of the pathway, we simulated pathways ranging from one to 100 consignments. Each consignment had 600 samples (30 units per crate and 20 crates) and its own

prevalence p_j sampled from a normal distribution with a mean = -5.29 (p=0.005 on the original scale) and a standard deviation σ=1 on the logit scale. We simulated the number of infested samples among crates within each consignment using a beta-binomial distribution with mean p_j and ρ=0.1. We repeated the simulation 10 times for each pathway size. We then fitted the hierarchical beta-binomial model to each replicated dataset to determine if we could recover the simulated parameters and see the uncertainties associated with these estimates.

While they are centred around their simulated value, one consignment with 600 samples is not enough to estimate p and ρ, and there is a large variation in the mean value of p and ρ among replicates. For both p and ρ, the standard error of the credible intervals is of the same size than the mean (coefficient of variation of 100%). This is not good enough. It seems that >30-40 consignments allow estimating ICC with sufficient precision (standard error of ~0.03, CV of ~0.3).

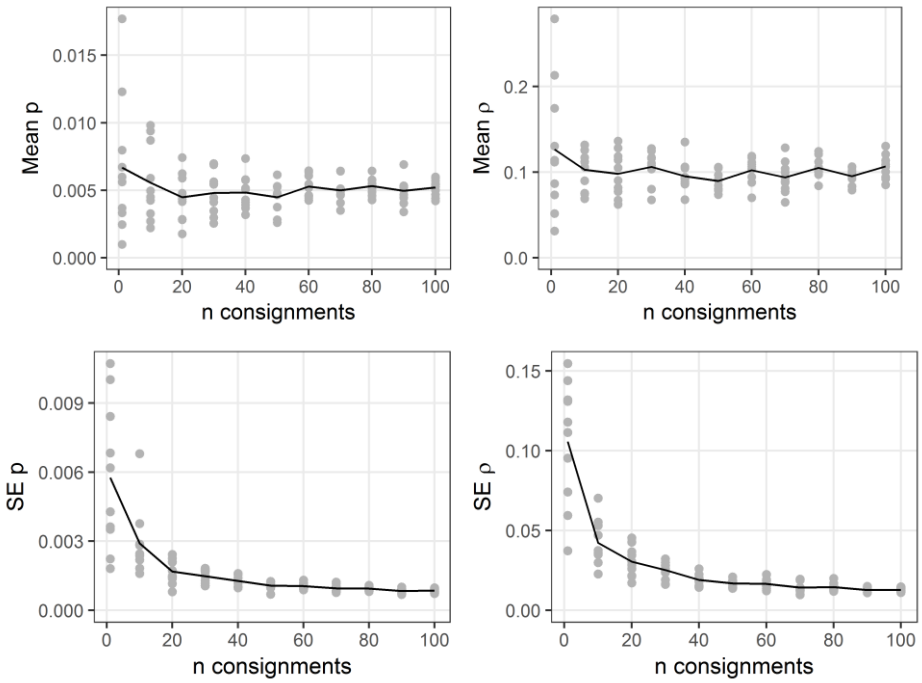


***Figure 6**: Mean parameter value and standard error for p and ρ estimated as a function of pathway size. Each grey dot shows one replicate simulation of the pathway, while solid black lines show the mean among different replicates.*

4.4.5. Comparison of hypergeometric and fixed proportion sampling results

An alternative to calculating sample size based on a desired inspection sensitivity is to do fixed proportion sampling. In biosecurity, it is common to sample 2% of the units contained in a lot/consignment. The sensitivity of proportion sampling is not constant and increases with lot size (**Table 3**). Inspection sensitivity can be particularly low when lot size is small (S=0.63 when N=10,000 and n=200). By contrast, hypergeometric sampling keeps the sensitivity constant, while the sample size increases with lot size.

***Table 3**: Comparison of the sensitivity (S) and the sample size (n) of proportional and hypergeometric sampling for varying consignment sizes (N). The infestation rate per consignment p was fixed at 0.005.*

	Proportion sampling		Hypergeometric sampling	
N	n	S	n	S
100	2	0.010	100	0.95
200	4	0.020	190	0.95
500	10	0.049	349	0.95
1,000	20	0.095	450	0.95
2,000	40	0.182	517	0.95
5,000	100	0.394	564	0.95
10,000	200	0.633	581	0.95
20,000	400	0.865	589	0.95
50,000	1,000	0.993	595	0.95
100,000	2,000	1.000	596	0.95
200,000	4,000	1.000	597	0.95

Slippage is the number of infested units not detected during inspection and that end up entering the country. Slippage is calculated as the proportion of consignments that are deemed compliant (1 – S) times the number of infested units in these consignments p(N-n).

For very small prevalence (e.g., p=0.0001), it takes many samples for fixed proportion and hypergeometric sampling to diverge in terms of leakage, as mentioned in the section 4.4.1.

Fixed proportional sampling is typically worse than hypergeometric sampling when prevalence and lot size are intermediate (p=0.001-0.005, N=10,000). Hypergeometric sampling is typically worse than fixed proportion sampling when prevalence is intermediate and lot size are large (p=0.001-0.005, N>50000). The leakage curves (Figure 7 and 8) intersect when the sample size *n* is the same for both sampling methods (~600 units, that is when N = 30,000). When sample size is high, hypergeometric sampling can do much worse than fixed proportion sampling (the leakage risk increases linearly with lot size).

One solution to this problem would be to use hypergeometric sampling when lot size is below 30,000 and use proportional sampling or increase the sensitivity of the inspection or the level of detection when lot size is higher as was done in Yamamura (1995, 2016).

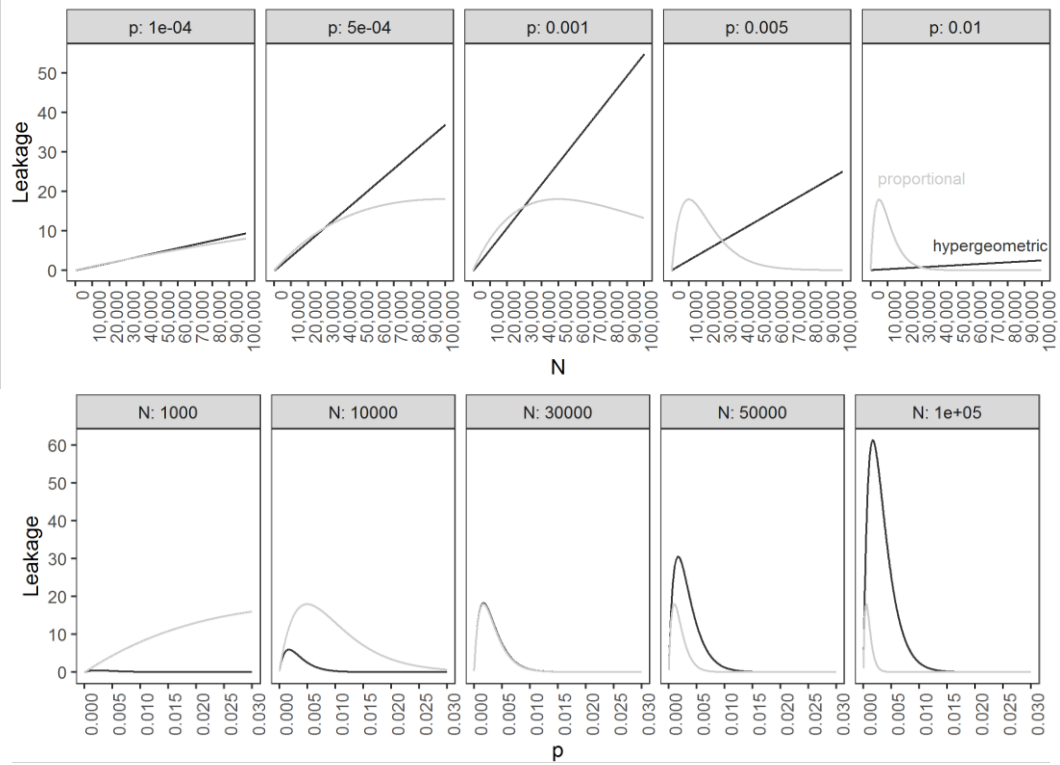


Figure 7: *Leakage of fixed proportion sampling and hypergeometric sampling as a function of lot size* N *and prevalence* p *in the lot. The sample size for the hypergeometric sampling was targeted to have a 95% confidence of detecting consignments having a prevalence of 0.005.*

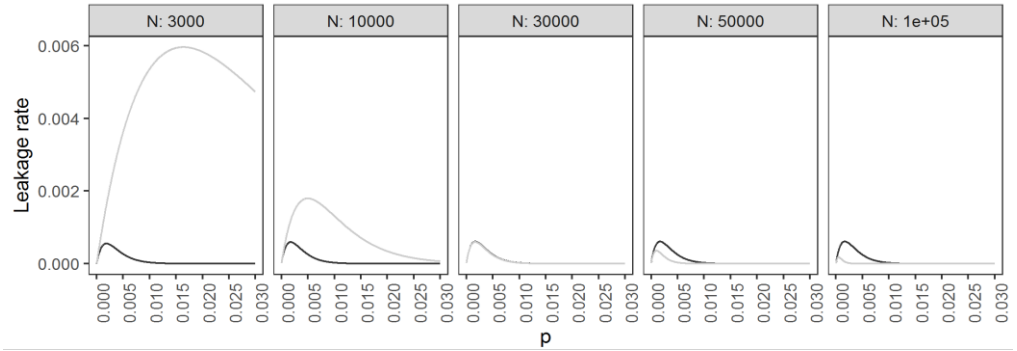


Figure 8: *Leakage rate (leakage / N) as a function of p. When N is low, leakage rate is much worst for fixed proportion than for hypergeometric sampling. Leakage rate for hypergeometric sampling is not sensitive to N. Another way to look at it is to determine the maximum leakage rate that we are willing to have and to set the sample size of the hypergeometric sampling so that the maximum of the leakage rate curve for any given p is below this value (Lane, et al. 2018.).*

4.5. Continuous Sampling Plans (CSPs)

The concept of continuous sampling plans was introduced by Dodge (1943) as sampling inspection plans for a product consisting of individual units manufactured in quantity by an essentially continuous process. The detailed procedure and tables for construction and selection of continuous sampling plans was provided by Stephens (1981). Bebbington *et al.* (2003) indicates that a CSP is a set of rules that provide a given Average Outgoing Quality (AOQ[1]), ideally with the minimum of effort (as measured by the Average Fraction Inspected or AFI). Most CSPs are based on the assumption that the quality (either defective or not) of successive production units is uncorrelated.

Continuous sampling plans may reduce inspection rates after an importer achieves a predetermined number of sequential pest-free consignments.

Different countries utilize CSPs in their plant health inspection procedures for acceptance of different products. Continuous sampling plans may reduce inspection rates after an importer achieves a predetermined number of sequential pest-free consignments. Examples of these applications will be provided below.

4.5.1. Different types of Continuous Sampling Plans

Dodge (1943) outlined the first type of continuous sampling plan or CSP-1. Almost a decade later Dodge and Torrey (1951) further refined CSP-1 and presented two additional continuous sampling plans - CSP-2 and CSP-3. (Antila, *et al.* 2008).

Schilling & Neubauer (2017) specified the procedure for CSP-1 for plant health purposes as follows-(Schilling and Neubaer 2017), also see **Figure 9**:

1. Specify the sampling fraction (***f***[2]) and clearing interval (***i***[3]) for the consignment.
2. Begin inspecting at 100%.
3. After ***i*** units in succession have been inspected and found without a defect proceed to randomly inspect the fraction (***f)*** of the units.
4. When a defective unit is found, revert to 100% inspection.

[1] Average Outgoing Quality, AOQ. The expected average quality level of the outgoing product, or mean fraction nonconforming in released lots, for a sampling plan for a given fraction nonconforming of the incoming product (Schilling and Neubauer, 2017) (also see section 5.2.2.2.).
[2] The f is the fraction of units to be tested, after the sampling has been triggered. For example, f = 1/10 means that every tenth product is tested.
[3] The parameter i, also known as clearing interval, defines the number of products needed to be tested fully (100%) before sampling can be started.

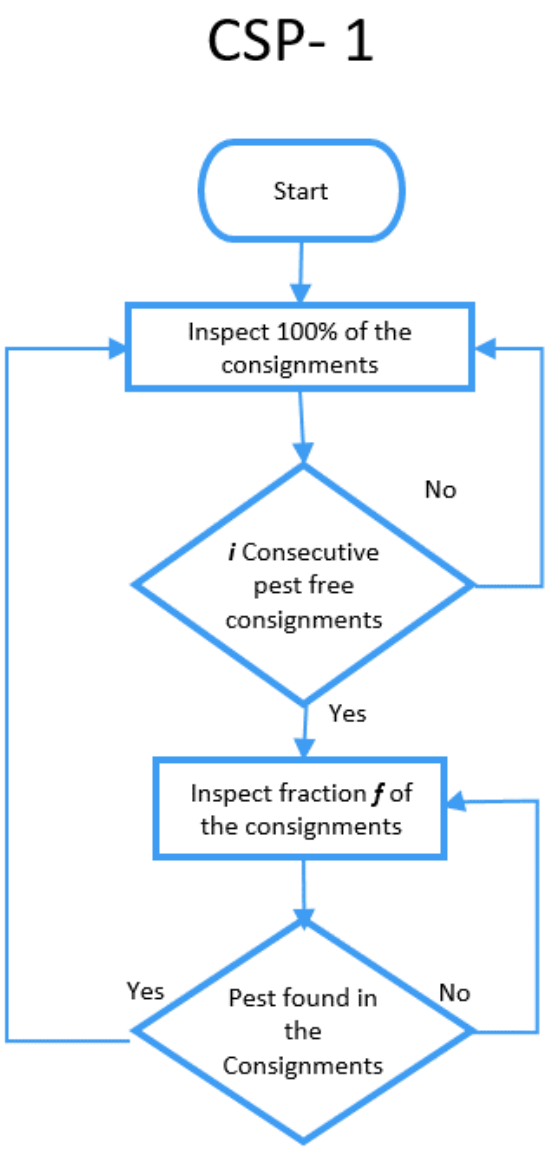


Figure 9. *Diagram for Continuous Sampling Plant 1 - CSP-1. Source: Adapted from (Antila, et al. 2008).*

Continuous Sampling Plan 2 - CSP-2 - is often preferred over CSP-1 as the return to 100% inspection does not occur immediately upon detection of a defective unit (Antila, *et al.* 2008). CSP-2 proposes to modify CSP-1 by changing some steps as follows:

When a defective unit is found, CSP-2 suggests that we continue sampling for ***k*** successive units. If no defect is found in these successive (***k)*** units, then we continue randomly inspecting a fraction (***f***) of the units. However, if a defective unit is found in the ***k*** samples, CSP-2 immediately reverts to 100% inspection (Schilling and Neubaer 2017).

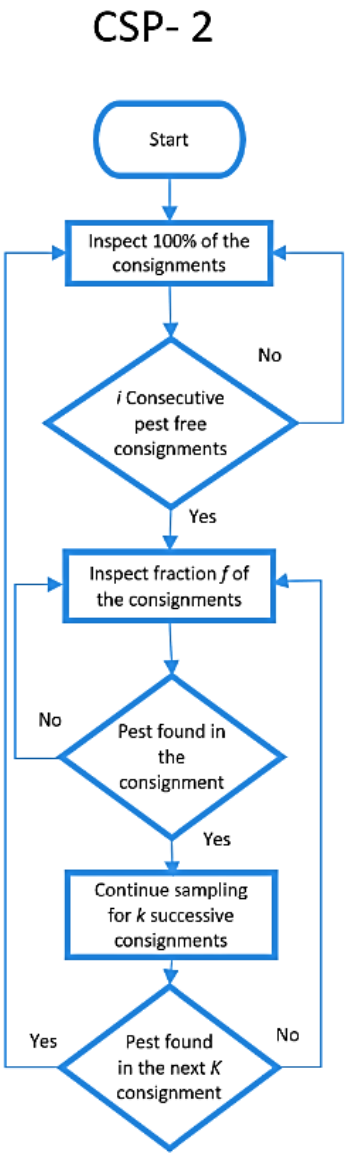


Figure 10. *Diagram for Continuous Sampling Plan 2 - CSP-2. Source: Adapted from (Antila, et al. 2008).*

In other words, the additional parameter, ***k***, is introduced in CSP-2. Compared to CSP-1, CSP-2 has the disadvantage of requiring a lengthened clearing interval *i* in order to reach the same AOQL[4] with the same fraction *f* (e.g., Juran, 1988; Dodge and Torrey, 1951; see also Antila*, et al.* 2008) (**Figure 10**).

CSP-3 introduces an additional refinement to CSP-1 and CSP-2. This refinement has the purpose of providing additional protection against variations in quality of the units sampled. In CSP-3, once the first defective unit is detected, the next four units are inspected. Should another defect

[4] Average Outgoing Quality Limit, AOQL. The maximum AOQ over possible values of fraction nonconforming for incoming products for a given acceptance sampling plan (Schilling and Neubauer, 2017) (also see section 5.2.2.2.).

be found among these four units, inspection reverts to 100%. However, if all four units are defect free, the CSP-3 plan continues as in CSP-2, and the next ***k*** units are sampled (see Juran, 1988; Dodge and Torrey, 1951; see also Antila, *et al.* 2008). See **Figure 11.**

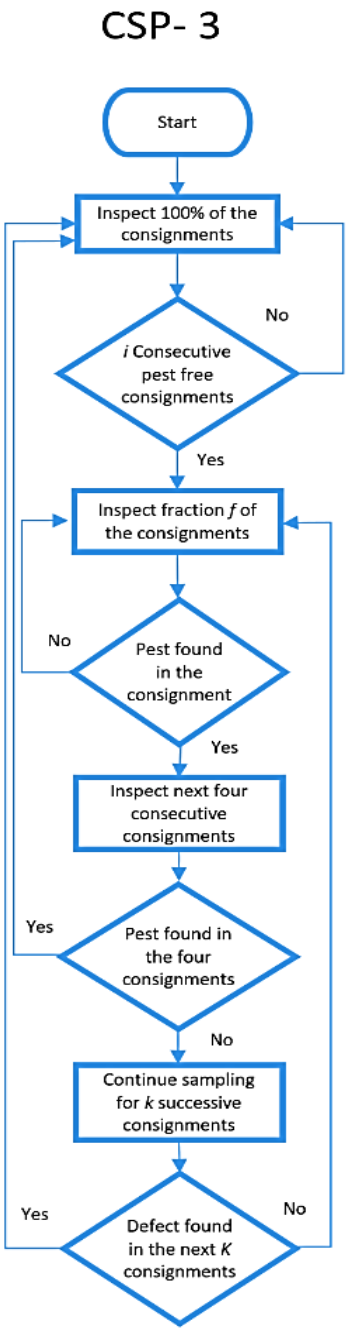


Figure 11. *Diagram for Continuous Sampling Plan 3 – CSP-3. Source: Adapted from (Antila, et al. 2008).*

4.5.2. Using the Continuous Sampling Plan for phytosanitary purposes

As indicated earlier, different plant protection or border inspection services use Continuous Sampling Plans in their inspection procedures for acceptance of different products. Below we provide some practical examples.

In the United States, the National Plant Protection Organization, USDA-APHIS Plant Protection and Quarantine (PPQ) utilizes different types of risk-based sampling plans, including ratings-based and continuous sampling plans. Ratings-based plans adjust inspections based on a commodity's analytically derived risk ranking. Continuous sampling plans reduce inspections after an importer achieves a predetermined number of sequential pest-free shipments (APHIS-USDA 2019).

These approaches to risk-based sampling were successfully tested during a 2-year a pilot at the PPQ Plant Inspection Stations, on plants for planting. They will be evaluated in the future in collaboration with the U.S. Department of Homeland Security Customs and Border Protection (CBP) for risk-based sampling of agricultural commodities arriving at U.S. ports of entry (APHIS-USDA 2019).

Australia has implemented the Compliance-Based Inspection Scheme (CBIS) through the application CSP methodology. CBIS uses historical data from selected pathways to reward consistently compliant importers through reduced inspections. This is an evidence-led and risk-based approach that allows the targeted re-allocation of inspection resources to higher risk pathways without compromising overall biosecurity outcomes.

According to the Australian Government, Department of Agriculture Water and Environment (2021), CBIS rewards importers of products who demonstrate consistent compliance with Australia's biosecurity requirements. Compliant importers benefit from the CBIS through smoother and more agile clearance of goods at the border and reduced regulatory costs. Once an importer has qualified under the CBIS, compliance of any future consignments will continue to be monitored under risk-based inspection rates at the line level and range from 10 to 50 % frequency. If a non-compliance is detected during the inspection or documentation assessment, the importer will return to 100 % inspection until they have once again demonstrated compliance and met the number of inspections required to re-qualify for the CBIS.

The inspection service in Mexico uses CSP-3 plans with protection levels of 95, 80 and 50%. Several manuals for inspecting seeds, grains, fruits and vegetables, dehydrated products, cut flowers and fresh foliage, and propagative plant material were developed by Ramírez Guzmán and López Tirado (2006 and 2007); see also Ramírez Guzmán 2017.

Analysis of the outcomes of implementing CSP-3 (Schilling, 1982) for low-risk importers in Mexico in 2013, demonstrated a 49.78% cost savings for inspection. The analysis assumed a CSP-3 scheme for importers with at least 3 years of zero quarantine pest detections (Ramírez Guzmán 2017).

Figure 12 show the characteristic operational curves for sampling frequency f=1/p, for f=0.20, 0.00418 and 0.001, with i=300 (i represents consignments that would be pest free after inspection at 100%) and k=4 (when a defective unit is found in 4 shipments, the next k successive units must be inspected, for this case k=4) for 95, 80 and 50% confidence, respectively, are shown.

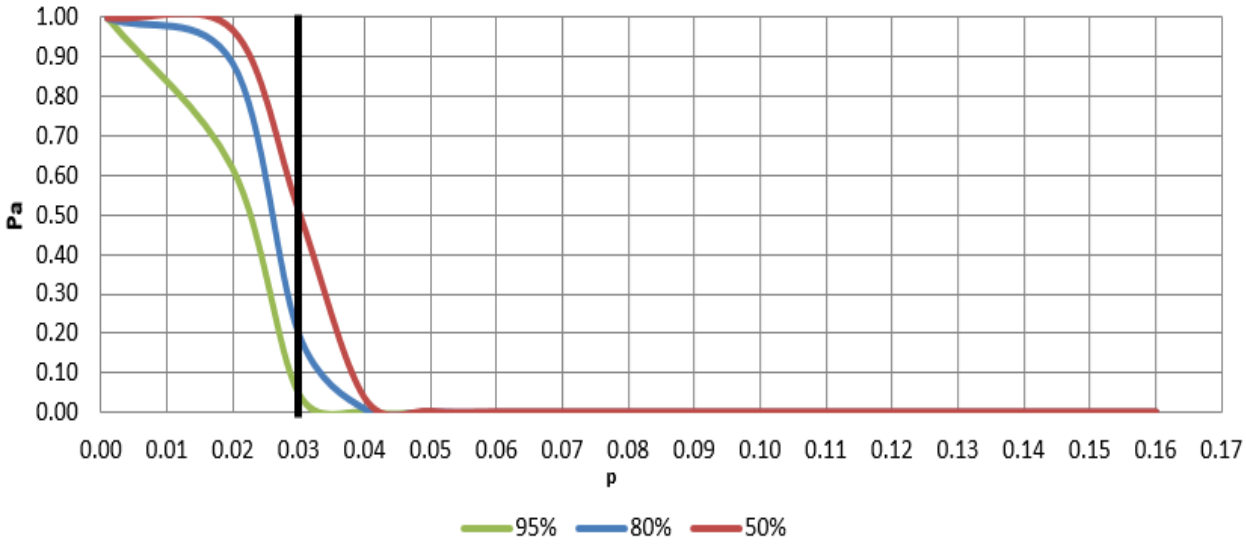


***Figure 12**. Characteristic operational curves for CSP-3 under different assumptions: 95%: i=300, f=0.02 and k=4, 80%: i=300, f=0.00418 and k=4 and 80%: i=300, f=0.001 and k=4.*

As can be observed (**Figure 12**), the three curves have a probability of rejection of 95, 80 and 50% at the height of p=0.03, which was what was expected according to Appendix H.

4.5.3. Example for CSP-3

Suppose that the NPPO of a particular country wishes to implement CSP-3 for a company that has presented up to 387 consignments without pests; from Appendix I we would expect that historical records would show, at most p1=0.015, which means that the proportion of consignments that have had positive pest finds has been at most 1.5%.

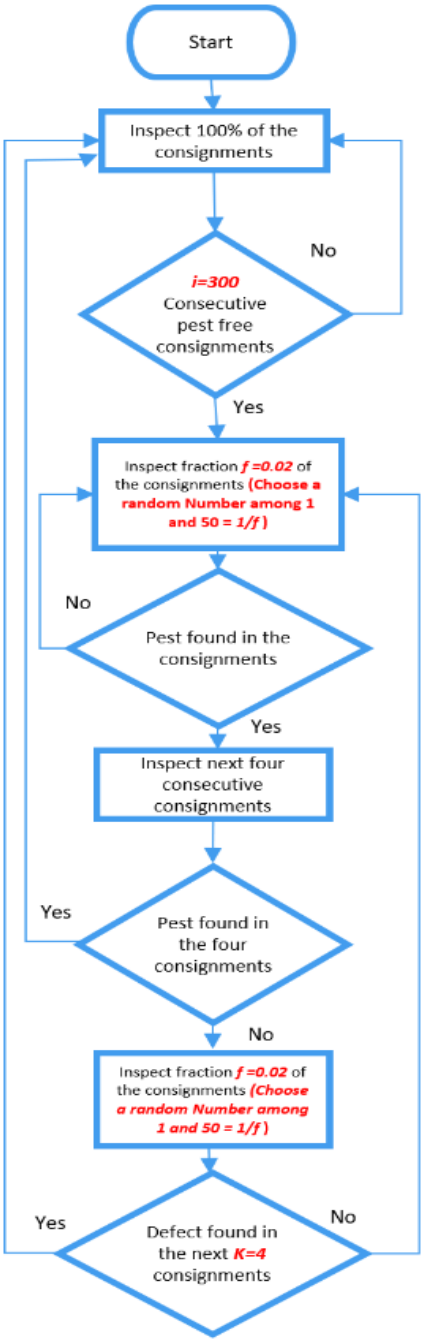


Figure 13. *Sampling plan CSP-3: f=0.02, i=300 y k=4.*

Following, one would select a sampling scheme from Appendix H. Let's suppose that the following sampling plan is selected (all plans where estimated with Monte Carlo simulation):

f=0.020, i=300 y k=4 (1)

This plan guarantees that if the process of generation of consignments begins to deteriorate, to the point of arriving at a subset of consignments with quarantine pests of p=0.03 (3%) this guarantees that consignments that do not meet with the original value of p=0.015 (1.5%) will be rejected with up to 95% probability. However, if one selects the plan:

$$f=0.00418,\ i=300\ y\ k=4 \qquad (2)$$

They will be rejected with 80% and with 50% probability if one works with the following (3) plan:

$$f=0.0010,\ i=300\ y\ k=4 \qquad (3)$$

Supposing that the plan (1) is selected, then the respective flow diagram would be the one shown in **Figure 13.**

As a conclusion, the application of CSP avoids 100% inspection of consignments of companies that for an acceptable period (for example three years) have consistently been free of quarantine pests. This strategy improves plant health, because the rejection of a consignment will incentivize the producer to take appropriate measures to ensure that his consignments are free of quarantine pests. This sampling methodology offers the opportunity for importing countries to make better decisions considering the proportion of consignments with quarantine pests, and the level of desired protection (Pr). In other words, using CSP sampling allows the selection of a sampling plan with a pre-determined level of confidence to reject consignments that do not meet the phytosanitary specifications.

Phytosanitary inspection (using a hand lens) of Hass avocado prior to export.
Source - https://www.senasa.gob.pe/senasacontigo/ica-inspeccion-fitosanitaria-de-palta-hass-para-exportacion-a-china/

5. DESIGNING, IMPLEMENTING, AND MAINTAINING A RISK-BASED SAMPLING PROGRAM

Barney Caton[1], Andrew Robinson[2]

1. Pest Exclusion Analysis Coordinator, Phytosanitary Advanced Analytics Team (PAAT). Center for Plant Health Science and Technology. Plant Protection and Quarantine, USDA
2. Director, CEBRA, School of Biosciences - Reader & Associate Professor in Applied Statistics Australia

5.1. Introduction

This chapter describes the processes that a plant or animal health protection organization can use to develop or refine a risk-based sampling (RBS) program for one or more pest pathways. The three major steps in that process are 1) design, 2) implementation, and 3) maintenance. Design refers to the development of the RBS program itself, or the identification of a suitable existing RBS program. Implementation is the process of preparing for and then initiating the RBS program on particular pathways. Maintenance describes the means by which the RBS program is monitored and adjusted over time to ensure that the objectives are being met and that safeguarding is effective.

Before beginning, we must agree that a **sampling plan** is needed to determine what sample size to select. An **RBS program** may include multiple sampling plans that are used depending on the history of inspection outcomes, or other risk factors. One of those sampling plans will always be normal inspection and additional plans will be either (i) reduced inspection, based on having met standards for higher product quality, or (ii) tightened inspection, based on having failed to meet certain standards. In this chapter we discuss how to create an RBS program including sampling plans for normal and reduced inspections and tightened inspection if desired.

The reasons for a National Plant Protection Organization (NPPO) to have an import inspection program include, in no particular order:

- gathering information about the pathway
- verifying compliance of imported goods (and make decisions about their enterability)
- intercepting potential plant health pests and diseases, or regulated goods that could carry them
- deterring malfeasance by producers and importers

For example, in Australia, the primary objective of inspections by the Department of Agriculture, Water and the Environment is to verify the compliance of the consignment with biosecurity regulations. Likewise, the NPPO of the United States has used sampling for detection for decades. More recently, agency experts agree that sampling for detection has limited phytosanitary and risk management value (see PPQ, 2016). Currently, the U.S. NPPO is shifting focus to value inspection for information and amplify the analytical potential of inspection results to improve and defend the role of exclusion in risk management (PPQ, 2016). This approach uses statistically designed randomized inspections. Although we focus on plant health biosecurity in this chapter, RBS programs can be used in exactly the same way for animal health protection (see Hood et al., 2019), or the safety of food and other products (see Mamber et al., 2018).

Why would an NPPO want to implement RBS? The trade- and standards-based rationales for using RBS were covered in the RBS manual Part I, as well as detailed descriptions of the advantages of using a more consistent procedure for sample size determination. In addition to those benefits, implementing an RBS program should directly improve inspection operations in two or more of the following ways:

1. Reduce the amount of (inspection) resources used on lower risk consignments
2. Provide incentives to producers and importers to improve the phytosanitary status of traded goods
3. Reallocate the resources needed for inspection of higher risk consignments
4. Reduce pests in the pathway or increase the number of pest detections in the pathway.

Goals 1 and 2 should occur as a consequence of designing and implementing an RBS program and may be what most NPPOs would like to achieve, while goal 3 must be explicitly built into an RBS program. Keep in mind that focusing only on goal 1 while maintaining normal sampling intensities on the rest of the consignments is likely to increase leakage of pests. Reducing leakage (goal 4) requires some producers and importers to increase conformity rates in response to incentives (i.e., reduced inspection). If an NPPO would like to achieve goal 4, then goal 3 should be a requirement for their RBS program design. This observation highlights the importance of assessing the performance of a program across multiple pathways, rather than in a piecemeal fashion.

The first step towards implementing an RBS program might be different for every NPPO. Because of this, we first present some characteristics that an agency's current inspection program should have and introduce newer features they may want to consider adopting. Next, we introduce two example, standard RBS programs, that are based on **cumulative results**. This type of approach uses recent inspection history to inform interventions at the border.

The bulk of the chapter describes the major steps to develop/refine a risk-based sampling (RBS) program for one or more pest pathways — namely, design, implementation, and maintenance. Throughout this section, we illustrate each step using the two standard RBS schemes in hypothetical case studies.

We note in passing that another type of RBS program is **ratings-based**, which is an approach that relies on statistical models of performance. At the end of the chapter, we briefly explain how ratings-based programs differ in design and discuss some advantages and disadvantages they have over cumulative results plans, including special considerations for model choice and ratings formulation.

5.2. Standard RBS programs – acceptance sampling

Sampling programs used to provide assurance that incoming commodities conform to some quality standard - like pest freedom - are generally known as **acceptance sampling** schemes (see Stephens, 2001). The objective of acceptance sampling for lots[5] is to ensure that producers submit lots that do not exceed an agreed-to level of nonconformities, which ensures that consumers receive lots that are acceptable (ISO, 2017). More simply stated, acceptance sampling facilitates the decision on whether to accept or reject a lot so that accepted lots conform to a standard (see ISO, 2013).

Acceptance sampling is used in many industries and military organizations, and it has been extensively studied since the early 1900s (Chen et al., 2017; Schilling and Neubauer, 2017). Statisticians and quality management experts have developed numerous acceptance sampling plans to address different inspection goals, pathways, or specific situations. Some pertain directly to agricultural imports, and we discuss below two pre-tested acceptance sampling plans as using these will be the simplest and most efficient way to implement an RBS program.

5.2.1. General concepts and definitions

Here we discuss some key concepts and definitions for RBS programs. Many of these have already been defined in Part I of the RBS manual.

Starting in section 5.2.4 below, this chapter follows a detailed discussion of two different RBS plans:

- MIL-STD-1916 and
- Skip-Lot Sampling

[5] In contrast, some acceptance sampling plans exist for inspection of units. Among the best known of these are 'continuous sampling plans' (also see section 4.5), which refer specifically to the inspection of individual units coming out of assembly line-type production processes (see Dodge 1943). These plans will not typically apply to the phytosanitary activities discussed here.

Both are **cumulative results** sampling plans, where future sampling levels are determined based on how many preceding lots have cleared without nonconformities since the last nonconformity was found in a lot (see Schilling and Neubauer, 2017).

In addition, both are sampling plans for **attributes** (see Stephens, 2001). This means we are sampling to answer a yes or no question about a particular attribute (for example, does the lot have regulated pests?). We might classify inspected items as conforming (pest-free, or free from other compliance problems) or non-conforming (infested with an actionable pest or having some other compliance issue). The alternative to attribute sampling is sampling for a **variable**, which involves some form of measurement (for example, how pure is this drug?).

In acceptance sampling, a distinction is made between a nonconforming lot and the items in the lot that caused the nonconformity. Such items are deemed to be **defective** so one can refer to some number of nonconforming lots with varying numbers of defective items. In phytosanitary terms, we make a distinction between actionable lots (or quarantine failures) and infested items (defective items).

Below are a few useful definitions, common specifications (parameters), or metrics used in acceptance sampling.

- **Acceptance quality limit, AQL**. The maximum percent nonconforming that can be considered satisfactory as a process average for the purposes of sampling inspection (ISO, 2013).
- **Acceptance number, *c***. The number of infested units or the number of individual pests that are permissible in a sample of a given size before phytosanitary action is taken (IPPC, 2008). For lots that comprise industrial products (e.g., components, parts, or 'widgets' generally), non-zero *c* values of nonconforming products in the sample may still indicate sufficient quality for that lot to be accepted (Stephens, 2001). In phytosanitary work, though, *c* will likely always be equal to zero, because allowing the entry of even one pest is deemed unacceptable. Plans with $c = 0$ are sometimes collectively called "zero acceptance plans."
- **Clearance number (Clearing interval), *i***. The number of successive lots that must be inspected and cleared with no nonconformities before a switching rule is invoked (Stephens, 2001).
- **Efficacy of detection**. The probability that an inspection or test of a defective unit(s) will detect the problem (IPPC, 2008). This is also called the sensitivity of inspection. In many cases this value will be less than 100 percent. We include this definition here for completeness, but we have not factored it into our examples below.
    - **Hypergeometric distribution.** A function that gives the probability of obtaining exactly *x* elements of one kind and $n - x$ elements of another if *n* elements are

chosen at random without replacement from a finite population containing *N* elements (see Roberts et al., 2015; Vose, 2000). This type of scenario is very common, such as in human epidemics, herd testing, and surveys.

- **Leakage (slippage)**. The number of accepted nonconforming items in a lot or a series of lots (Chen et al., 2017) (also see Glossary in RBS manual Part I).
- **Lot** (shipment). A number of units of a single commodity, identifiable by its homogeneity of composition, origin etc., forming part of a consignment (IPPC, 2008).
- **Lot size, N**. The total number of units of product in the lot (ISO, 2018). Lot size is controlled by the importer, not by the agency (EPPO, 2006) (also see Glossary in RBS manual Part I).
- **Operating characteristic curve (OC curve)**. Expected probability of acceptance for a lot or a series of lots under a given sampling plan (or scheme) as a function of the fraction nonconforming (quality level) (see **Figure 14** for an example) (Schilling and Neubauer, 2017; Stephens, 2001).

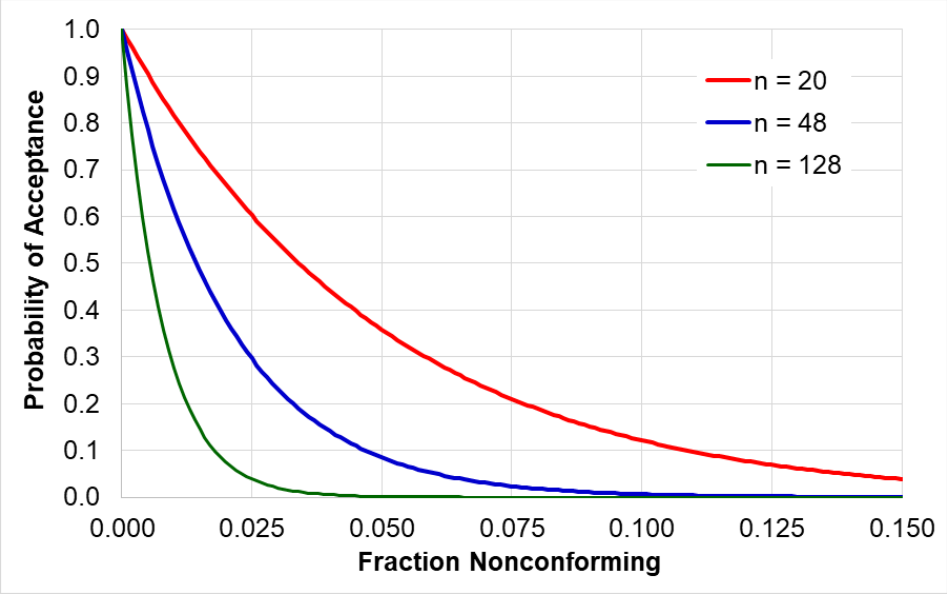


Figure 14. *Three operating characteristic curves, or probabilities of accepting a lot (N = 1,000) as a function of fraction nonconforming for three different sample sizes (n).*

- **Probability of acceptance, P_a**. The probability that a given sampling plan will accept (i.e., clear or not reject) lots at a given fraction nonconforming (or quality level).
- **Proportion of defective units, *d***. The mean fraction of defective units in a nonconforming lot.
- **Reduced inspection**. Decreased sampling, either in intensity or frequency, based on demonstrated product quality that has met set standards.
- **Reference sampling plan**. The normal (i.e., unaltered) procedure for determining sample size for a lot (after Stephens, 2001).

- **Sample size, *n***. The number of units selected from a lot that are inspected or tested (IPPC, 2008).
- **Switching rule**. An instruction within a sampling scheme for changing from one acceptance sampling plan to another of greater or lesser severity of sampling based on the demonstrated quality history (ISO, 2006). Typically, schemes have switching rules to move between normal, tightened, and reduced inspection plans.
- **Tightened inspection**. Increased sampling, either in intensity or frequency, based on demonstrated product quality that has exceeded standards set for normal inspection levels.

5.2.2. How much sampling to do?

Below we describe two different but related approaches to determine the appropriate sampling size. The first is based on the hypergeometric distribution (see **Table 4** below and Appendix 2 in RBS manual Part I). The second uses the Operating Characteristic (or OC) curve to assess impacts on the outgoing quality of lots (see **Figure 15** below). Both approaches are fundamentally hypergeometric processes.

5.2.2.1. Hypergeometric distribution approach.

This approach based on the hypergeometric function uses the following two parameters (Fosgate, 2009; IPPC, 2008) (Table 4; see Appendix for equations):

- **Acceptable nonconformity fraction, p_{Ref}**. The chosen (reference) minimum level for the proportion of incoming lots that are not acceptable (Stephens, 2001) (i.e., agencies wish to detect nonconformities at this level or greater). In phytosanitary terms, this is the **level of detection** or the proportion of lots infested with pests. Sometimes it is also referred to as quality level or pest action rate or, in hypergeometric tables, as the **acceptable risk level** (Anonymous, No date).
- **Confidence level, C_{Ref}**. The reference certainty of detecting a nonconformity at a given nonconforming fraction value (Daniel and Cross, 2013), or the probability that a lot with a fraction of defectives exceeding the level of detection will be detected (IPPC, 2008). Thus, 95 percent confidence means that, for lots with a given fraction nonconforming, 95 out of 100 would be detected.

Note that p_{Ref} and C_{Ref} provide information about the detection level for a single inspection. Determining how this would translate into the performance of a sampling plan over a number of lots would most likely require a simulation modeling approach. This helps to explain why standard acceptance sampling plans rely exclusively on the OC curve approach.

Finally, we distinguish between p_{Ref} and *p (proportion of lot)*. As indicated above, p_{Ref} is the estimated fraction nonconforming in lots subject to inspection, or the operational estimate of the quality of incoming lots. The estimated *p* could sometimes be greater than p_{Ref}, for low quality products. Also, the mean proportion defective, *d*, is different from p_{Ref}. It indicates the mean fraction of items in the lot that are defective.

***Table 4**. Partial reproduction of the hypergeometric sample size table for lot size = 1,000, based on confidence level and acceptable risk level, from Anonymous (No date).*

Acceptable risk level (p_{Ref})	Confidence level (C_{Ref})					
	0.8	0.85	0.9	0.95	0.99	0.999
0.0001	1,000	1,000	1,000	1,000	1,000	1,000
0.0002	1,000	1,000	1,000	1,000	1,000	1,000
0.0003	996	999	1,000	1,000	1,000	1,000
0.0004	983	992	998	1,000	1,000	1,000
0.0005	961	978	991	998	1,000	1,000
0.0006	932	958	979	994	1,000	1,000
0.0007	900	934	963	987	999	1,000
0.0008	867	907	944	977	997	1,000
0.0009	833	879	923	965	995	1,000
0.0010	800	850	900	950	990	999
0.0020	553	613	684	777	900	968
0.0030	415	469	536	631	784	900
0.0040	331	378	438	527	683	821
0.0050	275	316	369	450	601	748
0.0060	235	271	318	393	535	683
0.0070	205	237	280	348	481	626
0.0080	182	211	250	312	437	577
0.0090	164	190	225	282	399	534
0.0100	148	173	205	258	368	497
0.0200	77	90	108	138	204	290
0.0300	52	61	73	94	141	203
0.0400	39	46	55	71	107	156
0.0500	31	37	44	57	86	126
0.0600	26	31	37	48	72	106
0.0700	22	26	32	41	62	91
0.0800	20	23	28	36	54	80
0.0900	17	20	25	32	48	71
0.1000	16	18	22	29	43	64

5.2.2.2. Operating Characteristic (OC) curve approach.

In this approach, either or both of the following parameters could be used (Stephens, 2001):

- **Average Outgoing Quality, AOQ**. The expected average quality level of the outgoing product, or mean fraction nonconforming in released lots, for a sampling plan for a given fraction nonconforming of the incoming product (Schilling and Neubauer, 2017) (**Figure 15**). Because *AOQ* varies with *p*, which is unknown, summarizing it is not simple.
- **Average Outgoing Quality Limit, AOQL**. The maximum AOQ over possible values of fraction nonconforming for incoming products for a given acceptance sampling plan (Schilling and Neubauer, 2017) (**Figure 15**). This can be further defined on either a unit- or lot-basis:

  ***AOQL*$_1$** = unit AOQL, which is the upper bound on the long-run mean proportion of defective outgoing **product units** (on a sampling unit basis).

  ***AOQL*$_2$** = lot AOQL, which is the upper bound on the long-run mean proportion of nonconforming outgoing **lots**, or lots that would have failed the reference plan.

Some values for *AOQL* as a function of *n* are given in **Table 5.** The acceptance sampling reference books by Stephens (2001) and Schilling and Neubauer (2017) include several spreadsheet files for calculating these values for different sampling plans.

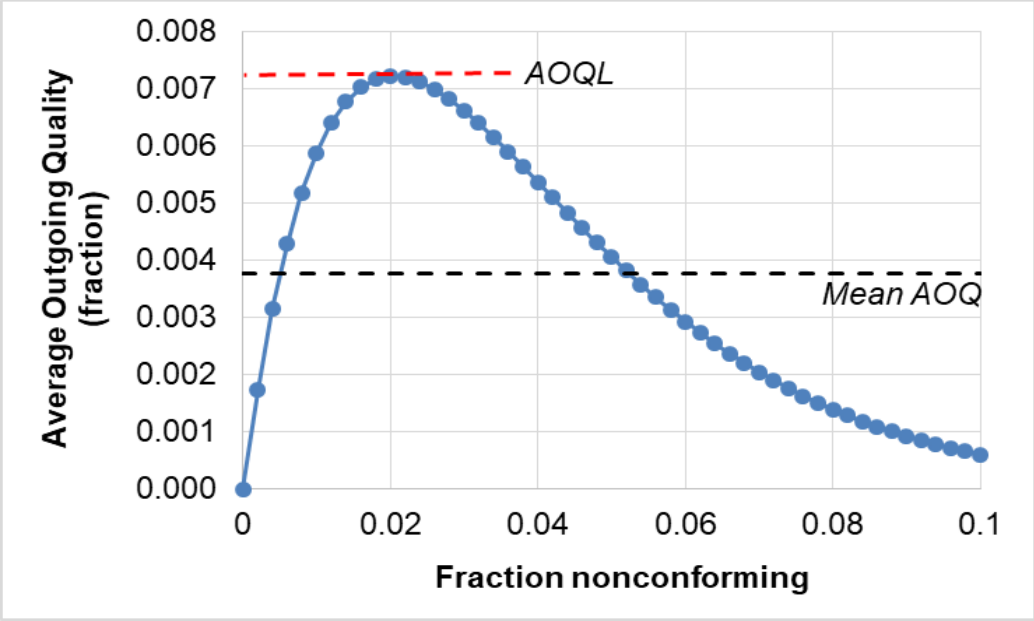


***Figure 15**. The average outgoing quality (AOQ; blue line and symbols) as a function of fraction nonconforming, with the average outgoing quality limit (AOQL) indicated by a dashed red line, and a dashed black line showing mean AOQ over the range from 0 to 0.1. Values shown are for lot size (N) = 1,000, and sample size (n) = 48.*

***Table 5**. Average Outgoing Quality Limit (AOQL) for given sample sizes, n, over different lot sizes, N.*

n	*AOQL*			
	Lot size (*N*)			
	500	1000	5000	10000
10	0.0361	0.0364	0.0367	0.0368
20	0.0177	0.0180	0.0183	0.0184
30	0.0115	0.0119	0.0122	0.0122
40	0.00846	0.00883	0.00912	0.00916
50	0.00662	0.00699	0.00728	0.00732
60	0.00540	0.00576	0.00606	0.00609
70	0.00452	0.00489	0.00518	0.00522
80	0.00386	0.00423	0.00453	0.00456
90	0.00335	0.00372	0.00401	0.00405
100	0.00294	0.00331	0.00361	0.00364
110	0.00261	0.00298	0.00327	0.00331
120	0.00233	0.00270	0.00299	0.00303
130	0.00209	0.00246	0.00276	0.00279
140	0.00189	0.00226	0.00255	0.00259
150	0.00172	0.00208	0.00238	0.00242
160	0.00156	0.00193	0.00223	0.00226
170	0.00143	0.00180	0.00209	0.00213
180	0.00131	0.00168	0.00197	0.00201
190	0.00120	0.00157	0.00186	0.00190
200	0.00110	0.00147	0.00177	0.00180
210	0.00102	0.00138	0.00168	0.00172
220	0.000936	0.00130	0.00160	0.00164
230	0.000864	0.00123	0.00153	0.00156
240	0.000797	0.00117	0.00146	0.00150
250	0.000736	0.00110	0.00140	0.00143
260	0.000679	0.00105	0.00134	0.00138
270	0.000627	0.000995	0.00129	0.00133
280	0.000578	0.000946	0.00124	0.00128
290	0.000533	0.000901	0.00120	0.00123
300	0.000491	0.000858	0.00115	0.00119

5.2.3. Examples

Here we demonstrate how each approach mentioned above can be used to identify the level of detection for particular sampling plans, report the impact on the needed effort, and indicate how they relate to one another. In all cases the lot size (*N*) is 1,000 units.

Example 1. If the NPPO wishes to have a confidence level (C_{Ref}) of 95 percent and an acceptable risk level (p_{Ref}) of 6 percent (0.06), using the hypergeometric function approach (**Table 4**), then we determine that the number of samples needed for the lot size of 1,000 is 48 samples.

Working backwards from *n*, we can find the equivalent using the OC curve approach. We include this here to facilitate comparison. Using the OC curve approach, *n* = 48 gives an $AOQL_1$ of 0.0073 (0.73 percent) [see Eqn. A5 in Appendix A].

Example 2. If the NPPO wishes to have a mean *AOQL* of 0.001, then *n* is approximately 265 (**Table 5**). [In more exact terms, *n* = 268.]

Again, working backwards from *n*, in terms of the hypergeometric function, *n* = 265 corresponds to multiple combinations of C_{Ref} and p_{Ref} , but if we choose C_{Ref} = 0.95, then p_{Ref} is approximately 0.0097 (**Table 4**), or about 1 percent. As sample size increases, the level of detection achieved also increases.

5.2.4. Two examples of standard RBS plans

The chapter will use two examples of RBS plans that can be applied to pathways or lots. In MIL-STD-1916, every lot is inspected but with differing degrees of intensity. In Skip-Lot Sampling, as the name implies, the inspection effort is reduced by clearing some lots without inspection, but all remaining lots are inspected with the same degree of intensity.

5.2.4.1. MIL-STD-1916 - all lots inspected; some with reduced sampling intensity

This approach consists of three sets of matched sampling plans for lots, indexed by seven different verification levels, or quality standards (Department of Defense, 1996).[6] The three sampling plans are **normal, reduced (or lowered)** and **tightened**, with adjustments affecting sample size or sampling intensity (i.e., every lot is inspected).

Sample sizes are determined using a table (reproduced below as **Table 6**) based on:

1. The desired verification level (baseline level of detection)
2. The reference sampling plan
3. Lot size - indicated by code letters (A-E).

The switching rules between the three plans is as follows:

- *Normal to tightened* - 2 nonconforming lots have been found in the last 5 or fewer lots

[6] Note that the approach also contains a continuous sampling plan section, but this is for application to units and not lots, and so is irrelevant to most phytosanitary inspections.

- *Tightened to normal* - 5 consecutive lots have been accepted (ideally with some indication of the quality issue was rectified)
- *Normal to lowered* - 10 consecutive lots have been accepted while on normal inspection (and assuming regular volumes of largely homogenous products)
- *Lowered to normal* - nonconforming lot has been found (or, perhaps, volume becomes irregular or stops, or other information indicates production problems or quality issues)

Beyond that, a new sampling plan for a change towards tightened inspection (i.e., a larger sample size) is simply the next higher verification level in **Table 6**, or, for a change towards reduced inspection (i.e., a smaller sample size), the next lower verification level in **Table 6** (Department of Defense, 1996). In summary, **Table 6** describes all seven sampling plans which might be chosen.

To illustrate, the 48-unit sample for a lot size of 1,000 applied to MIL-STD-1916 corresponds to verification level III (reference sampling plan) (**Table 6**). Tightened inspection (verification level IV) for the same lot size gives a sample size of 128, while a lowered inspection (verification level II) gives a sample size of 20 (**Table 6**).[7] In this example, the inspection savings from lowered inspection with MIL-STD-1916 is from examining 28 fewer units per lot but continuing to inspect every lot.

Table 6. *Reproduction of Table II in MIL-STD-1916 for determining sample size based on the verification level (T-R) and lot size (Code letter) (Department of Defense, 1996). The sampling plan for lowered inspection is one verification level to the right of the default verification level chosen (reference sampling plan), while that for tightened inspection is one verification level to the left.*

Code letter	Verification Levels								
	T	VII	VI	V	IV	III	II	I	R
A	3072	1280	512	192	80	32	12	5	3
B	4096	1536	640	256	96	40	16	6	3
C	5120	2048	768	320	128	48	20	8	3
D	6144	2560	1024	384	160	64	24	10	4
E	8192	3072	1280	512	192	80	32	12	5

5.2.4.2. Skip-lot sampling - not all lots inspected; consistent inspection intensity among lots

Skip-lot sampling[8] is defined as a "*sampling inspection procedure in which some lots in a series are accepted without inspection when the sampling results for a stated number of immediately preceding lots meet stated criteria*" (ISO, 2005). The purpose of skip-lot sampling "*is to provide a*

[7] In alternative terms, a sample size of 148 equates to 95 percent confidence in finding a fraction nonconforming of 2 percent, while a sample size of 20 is 95 percent confidence in finding a fraction nonconforming of 14 percent.
[8] This should not be confused with continuous sampling plans, which are similar approaches but applied to individual units—not lots or batches—coming from an assembly-line-like process (ISO, 2005). (See Continuous Sampling Plans (CSPs) in section 4.5)

way of reducing the inspection effort on products of high quality submitted by a supplier who has a satisfactory quality assurance system and effective quality controls". Often, NPPOs may substitute knowledge that agricultural production and processing tend to reduce pest presence in products as a proxy for direct, documented information about producers' or shippers' quality assurance systems and controls.

Reduced inspection is achieved by randomly determining, at a specified probability, if an incoming lot that qualifies for reduced inspection will be accepted without inspection. Thus, the unique parameter to be set for skip-lot sampling is the **inspection frequency** (***f***) which is the proportion of lots that will be sampled (on average) using the reference sampling plan when lot skipping is active (Stephens, 2001; ISO, 2005). Because the reference sampling plan is used for every inspection, data collection in skip-lot sampling approaches is more consistent than in reduced intensity approaches. Moving forward, we will refer to the reduced inspection level(s) in this scheme as 'skipping inspection,' and to normal inspection as qualifying inspection.

ISO standard 2859-3 specifies that qualification for skipping inspection uses a constant clearance interval, $i = 10$, as well as a second standard, called the qualification score (see ISO, 2005). It also sets the minimum achievable value of $f = 0.2$, stepping down through f values of 0.5, 0.4, and 0.3.[9]

In our examples we will use the more flexible approach described as SkSP-2 (Skip-Lot Sampling Plan 2) by, for example, Stephens (2001) and Schilling and Neubauer (2017). Using SkSP-2, one can devise specialized inspection approaches that meet the desired *AQL* or *AOQL* (**Table 7**), and that could use values of i other than 10, and perhaps f values below 0.2, if deemed appropriate.

The switching rules are as follows:[10]

- *Qualifying to skipping*: i consecutive lots are accepted (i.e., no nonconformities) while on normal inspection
- *Skipping to qualifying*: A nonconforming lot is found

While the interval when the consecutive inspections should occur is not technically specified, the general guideline, as in MIL-STD-1916, is that shipping volume should be fairly continuous. Agencies could adopt procedures to reset inspection levels after periods of inactivity or set a minimum number of lots per particular time, if they desire.

[9] The ISO plan also recommends against using $c = 0$ plans, reportedly because they have "poor switching characteristics" compared to plans with $c \geq 1$. For arguments in favor of $c = 0$, see Department of Defense, 1999).
[10] A 'tightened' option is typically not included in these schemes, since increasing f above 1 is not possible. Such an option might be built by increasing sampling intensity (altering the reference plan) under specified circumstances.

As an example, a SkSP-2 plan could specify i = 10 and f = 0.4, with a second level of f = 0.1 reached after a further i = 10 acceptances. Using the example reference sampling plan from above (C_{Ref} = 0.95, p_{Ref} = 0.06; or *AOQL* = 0.0073), *every* inspection for N = 1,000 will have n = 48 samples.

A product with 10 consecutively accepted lots would change from qualifying to skipping inspection. At level one, with f = 0.4, each lot would have a 4-in-10 chance of being inspected, or, equivalently, a 6-in-10 chance of being accepted without inspection.[11] The corresponding *AOQL* for that level of inspection is 0.035 (**Table 7**). The increase in *AOQL* under SkSP-2 compared to the *AOQL* for the reference plan reflects leakage because of skipping. After another 10 inspected lots have been accepted, the product would move to the second skipping level, with only a 1-in-10 chance for inspection. That level of inspection has an *AOQL* of 0.052 (i = 10 + 10 = 20; **Table 7**).

***Table 7**. Approximate values for the Average Outgoing Quality Limit, AOQL, for skip-lot sampling plans with different clearance numbers (i) and inspection frequencies (f), based on equations in Stephens (2001).*

Clearance number, i	Inspection frequency, f				
	0.5	0.4	0.3	0.2	0.1
8	0.032	0.043	0.058	0.081	0.121
10	0.026	0.035	0.048	0.065	0.099
12	0.022	0.029	0.039	0.056	0.083
14	0.019	0.026	0.035	0.047	0.073
16	0.016	0.022	0.031	0.042	0.064
18	0.014	0.019	0.027	0.038	0.056
20	0.013	0.017	0.023	0.034	0.052

5.2.4.3. Other sampling plans

Numerous other standard plans exist and could be used. Good sources of information, particularly about zero acceptance sampling plans (see 'acceptance number,' above), include the following - Shmueli, 2016; Squeglia, 2008; Stephens, 1995.

[11] Note that probabilities should be used as such. For products in the skipping inspection phase, every time a lot arrives, inspection vs. acceptance should be determined randomly. In other words, a 1-in-10 chance (f = 0.1) must not be operationalized as "Inspect the first lot and skip the next nine." The average inspection frequency over time will equal f, but over short intervals it could deviate substantially (e.g., 3 or 4 lots in a row selected for inspection). The ISO 2859 standard has an appendix with practical descriptions of how agencies can randomize selections (ISO, 2005).

5.3. Assessing current sampling operations

5.3.1. Basic features of an inspection program

Before an NPPO should consider adopting an RBS program, their existing inspection program needs need to be identified and understood. This is preparation (i.e., step 0) for implementing an RBS program.

5.3.1.1. Existing inspection programs

An NPPO should already be regularly conducting import inspections on the pathway(s) of interest. Trained personnel and infrastructure for inspection should be in place, as well as mechanisms for data and shipment processing. Implementation of an RBS program will be easier if the NPPO is already using inspection data for targeting or monitoring the pathway.

Inspection is never 100 percent effective (see Yamamura et al., 2016). Because of this, NPPOs expect that the cumulative effect of all phytosanitary measures, including inspection, will reduce propagule pressure to an acceptable level of protection. This helps ensure that any problematic pest or disease organisms that gains entry into a country is isolated and rare enough that a population will not be able to establish (survive and reproduce) (Blackburn et al., 2011). The assemblage of risk mitigating actions has been called the biosecurity (Beale et al., 2008) or safeguarding continuum (see PPQ, 2015), and it is important to note that import inspections play only a small part (**Figure 16**) in the continuum.

In addition to having an inspection program in place, the pathway(s) of interest should have a reasonable volume of consignments. Without substantial volumes of goods to inspect, substantial numbers of entities/commodities may not qualify for reduced inspections, due to a lack of statistical significance or not reaching cumulative thresholds. In that case, the benefits of adopting RBS programs may not be worth the effort required for implementation and maintenance. The meaning of reasonable volume may vary, but would be met if, for example, the number of lots inspected each year numbered in the upper hundreds or thousands.

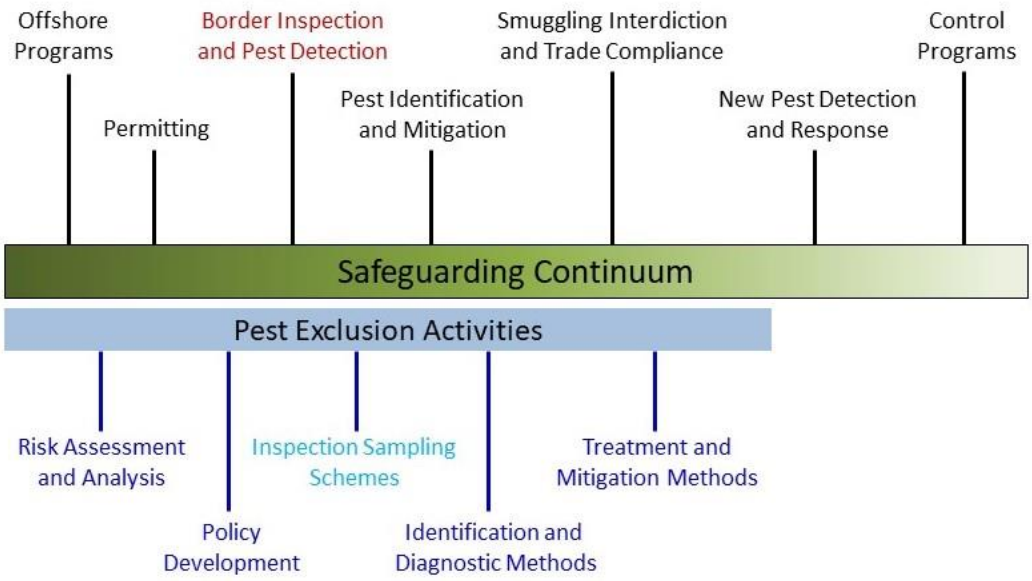


***Figure 16**. Schematic of the Safeguarding Continuum by which multiple measures and actions are used to manage the risk of introduction and spread of non-native pests and diseases.*

5.3.1.2. Reference sampling plans

A reference sampling plan is used for choosing the sample size to inspect a routine lot. The sample size needs to be specified, because an RBS scheme may adjust the reference sampling plan in some way (e.g., reduced sampling intensity).

Reference plans may be defined within a standard plan or may be customized. An example of a standard-based reference plan will be presented below. For a customized example, in Australia many plant product inspections conducted by the Dept. of Agriculture and Water Resources (DAWR) use a hypergeometric table to calculate the sample size based on a 95 percent confidence of finding a 0.5 percent infestation (Robinson, 2018). Hypergeometric sample sizes depend on the size of the lot, but for a lot size of 1,000 that customized reference sampling plan gives a sample size of 450 (**Table 4**).

In contrast, the U.S. Dept. of Agriculture, uses a reference sampling plan based on 95 percent confidence in detecting a 5 percent infestation rate. Under this plan, a lot size of 1,000 would require a sample size of 57 (**Table 4**).

5.3.1.3. Incoming consignment information

Most NPPOs collect information about incoming consignments in order to document possible pest pathways and capturing information for more efficient inspection schemes is a secondary concern. But in an RBS program information becomes a primary concern because inspectional status depends explicitly on historical information about the commodity and, perhaps, about the

entities involved. Here we assume that details concerning the plant products, including name (scientific or common), quantity and units of measure, and country of origin are routinely recorded by NPPOs. When the United States initiates trade in a new plant commodity, the information requirements include the "Globally Unique Product Identification Code" (APHIS, 2016).

NPPOs also need to identify and collect any additional information required for their RBS program. Formalizing the collection of information about producers or importers is one possible example. Once identified, the information might be a mandatory data requirement from importers of those products. The data collection system may need to be enhanced/modified to accommodate new data field(s). For example, the USDA-APHIS-PPQ recently started collecting data on 'type of propagative material' (e.g., rooted plant, unrooted cutting) imported, which is useful for its RBS program.

5.3.1.4. Recording inspection results

Here we provide an overview of the commonly recorded data related to inspection outcomes. NPPOs may or may not choose to record this data, depending on objectives and activities. A possible template for collection of some useful inspection data can be seen in section 3.1, above.

Inspection outcomes. This is the most basic/critical piece of information to be collected during inspection: that is, whether or not the lot or commodity was infested by a pest of concern. Counting the number of pest actions in recently inspected lots is a common and mathematically straightforward way of determining inspection levels in a number of standard RBS plans.

It is important to note that all detected pests may not be equally relevant (see Yamamura et al., 2016). Pests can be (i) those that are "in or on" the commodity, or known to use the commodity as a host, or can be (ii) hitchhikers (contaminating pests), that is, those that may be associated with the consignment, conveyance, or packaging material, or present as a consequence of packing conditions in the exporting country (e.g., packing under artificial light). Whereas pests on the commodity are likely to be relevant to an RBS program, NPPOs will determine whether or not contaminating pests count toward program status determinations.

Detections of actionable (= those that result in some regulatory action by the NPPO) pests can be recorded in different ways. For example, the datasheet can include a 'pest action flag' [the name is customizable], which is equal to 1 if one or more actionable pests were found in the lot, or 0 if no pests were found.

Actions taken. Actionable pests may not necessarily be handled the same way. For example, the USDA-APHIS-PPQ uses disposition codes to indicate the type of pest-related actions taken (e.g., fumigated or re-exported) as well as other actions related to consignment compliance (e.g.,

prohibited commodity, missing phytosanitary certification) (PPQ, 2018b). This information may be important to record for understanding trends and impacts on inspection operations, such as for estimating the time spent on fumigating or destroying infested lots.

Pest information. Information on the taxonomy, life stage(s) and number of pests found helps NPPOs to understand the level of risk posed by pests in the pathway. The information may provide feedback to producers and shippers that want to improve commodity/consignment quality. Accurate pest identification can be challenging (see Floyd et al., 2010), because there are many thousands of plant pest species, one can detect life stages that are difficult to accurately identify (e.g., arthropod eggs), and because timely decisions on pest actions are needed for perishable commodities. Before considering a shift to an RBS program, NPPOs should ensure that they have reliable and rapid pest identification expertise relative to the volume of trade they experience.

Entities. Plant product import supply chains begin with producers and suppliers in the exporting country, followed by import inspections, and finally with consignees/receivers in the importing country (**Figure 17**).[12] It may be challenging to record data on each entity involved in a particular pathway. Nevertheless, NPPOs should attempt to record information for the entity level at which they expect to manage the RBS program (see 5.5.1.2). Pest presence and their management in crops seem likely to be more closely linked to producers than to suppliers and importers, but this is not certain (Griffin [NAPPO], personal communication). Thus, while it may seem ideal to record information on producers to manage the RBS program (e.g., feedback processes), in practice information may be more readily available for suppliers and importers. In addition, if the pathway includes many small-volume producers, aggregating outcomes at higher levels may be preferable, to increase potential eligibility for inspection reductions. An advantage of recording information at the level of importer may be simpler communications because they may often be located in-country. A disadvantage may be that although they have a choice of suppliers, importers are somewhat removed from production and mitigation operations and may have limited ability to affect management practices.

Recording information about entities will be greatly facilitated in the coming years as the WTO-Trade Facilitation Agreement (WTO-TF) begins requiring electronic consignment data, via the single window system (ECFE, 2005).

Plan-specific information. Some information specific to the inspection scheme may need to be collected. This might include required codes or identifying numbers particularly if only some

[12] Export inspections may also be done by the exporting NPPO to produce phytosanitary certificates, but that agency is not part of the supply chain for the purposes of an RBS program. In other words, it is unlikely that an RBS program would be administered by monitoring the exporting NPPOs.

commodities are eligible for and processed through the RBS program and such codes are used as identifiers.

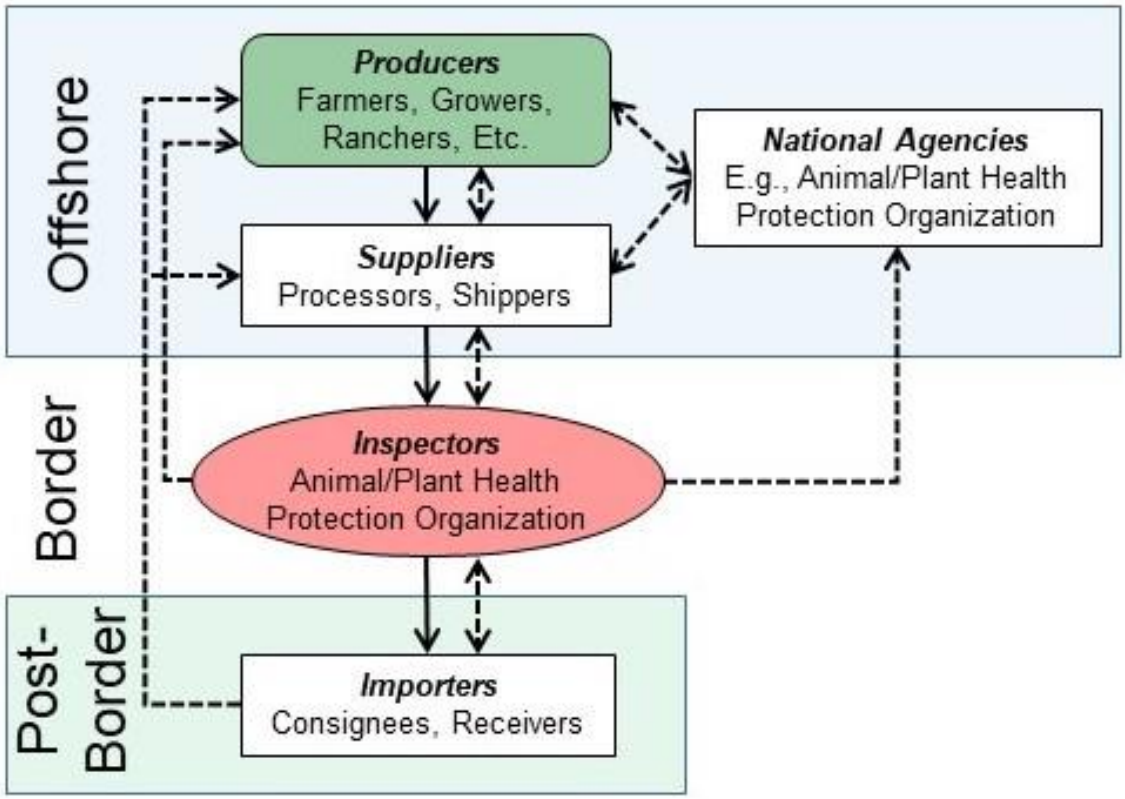


Figure 17*. An import supply chain for plant products, including inspection at the border. Solid lines represent material transfers while dashed lines are possible flows of information.*

5.3.1.5. Recording recent inspection results

Most acceptance sampling plans will work best when (i) the pathway volume is reasonably consistent and high (hundreds of lots annually, rather than, say, dozens), (ii) producers and shippers are known to be fairly consistent in expected commodity quality (either good or bad), and (iii) commodities and their packaging are reasonably homogenous (Stephens, 2001). Consistency helps ensure the relevancy of recent results and maintenance of production practices, while high volumes confer greater potential eligibility for reduced inspections. Understanding the import histories of relevant entities will inform the design process and estimation of potential impacts. Managing homogenous products could allow more specialized programs to be developed and helps ensure that the inspection scheme will be both appropriate and manageable.

Therefore, we suggest that the amount of data needed should be that which is sufficient to convince the NPPO that the above standards have been met. Later we show how to use such

data to estimate the impact the new RBS plan would have on overall inspection resources and safeguarding outcomes (see 5.5.4). For commodities with volumes that are not highly seasonal, six months of previous inspection results may be adequate to estimate operation variables for a proposed new RBS program. Current estimates of leakage would be helpful in determining current program efficiency, and for comparing it to estimated outcomes after implementation.

5.3.2. Additional features for more effective RBS programs

5.3.2.1. Dynamically updated inspection levels

In addition to the features mentioned above that should be in place before an NPPO decides to implement RBS, there is an important feature that NPPOs need to plan create, preferably before, or at least concurrently with implementation of RBS. This feature, within data management system(s), would (i) automatically determine whether incoming lots are inspected at normal or reduced (or tightened) levels, (ii) display this information as needed, and if possible (iii) directly determine and display the sampling protocol to be used. In general, the larger the volume of imports and products in the RBS program, the more automated the data system will need to be. If thousands of entities or commodities are being tracked, then updating inspection status manually will be very resource-intensive, and prone to errors, which will undermine confidence.

Proceeding in a stepwise fashion, the basic data systems need to display the inspection results to assess recent outcomes and determine how many of the last (sequential) *i* lots were nonconforming. This data would be used to assign an inspection level to the commodity (i.e., invoke switching rules as needed). This involves a simple counting exercise. The challenges will be in having access to recent inspection data, assessing the correct series of consignments, and determining their status.

5.3.2.2. Dynamic sampling inspection support

The next optional steps are to (i) determine the effect of the level of inspection on the sampling plan, (ii) calculate the sample size, and (iii) randomly select the samples. Plant products subjected to inspection will require the reference sampling plan, but reduced or tightened inspection levels should be subject to a different plan, or to no sampling plan if inspection will be skipped. Procedural options by level of complexity (i.e., how much is done in the system vs. outside the system) are shown below (**Table 8**).

***Table 8**. Comparison of data system tasks and tasks for inspectors with increasing automation of the data system. The level of complexity goes from simple to complex, and the system tasks are cumulative and include all tasks above the selected line.*

Level of complexity	Cumulative system tasks	Inspector tasks
Simple	Determine and display the inspection level	Determine which sampling plan is followed and carry out all inspection procedures, perhaps using work instructions
	AND display the associated sampling plan	Carry out all remaining inspection procedures, starting with random determination of inspection if warranted
Moderately complex	AND if inspection is a possibility, randomly determine whether to inspect the lot	Carry out all remaining inspection procedures, starting with sample size determination
	AND use information about the lot to calculate the sample size	Carry out all remaining inspection procedures, starting with randomly selecting samples
Complex	AND use lot information to randomly determine which samples to select	Remove targeted samples from the lot, inspect and report results

None of the functions indicated above is complicated, even at the highest level of complexity; most are straightforward arithmetic functions. Randomization processes are often built into standard code systems, and sample sizes extracted from a standard table could use a lookup function on the master table (or tables) in the system, which can be updated, as necessary. As more sampling procedures are built into the system, the number of tasks performed by inspectors decreases. At full automation, inspectors only have to select the proper samples and carry out the inspection. This has obvious benefits for reducing errors and correctly selecting the random samples (for more on inspection errors, see Collins et al., 1973; Minton, 1972).

5.3.2.3. Other considerations for dynamic information systems

The considerations below refer to operational factors that each NPPO needs to address.

Timing of updates. Ideally, the data system would operate in real-time, continuously updating inspection levels as inspection outcomes become available. For NPPOs dealing with large consignment volumes arriving at different ports-of-entry, a daily update may be sufficient. Longer delays between updates negatively impact RBS program responsiveness to changes in pest action rates and may undermine stakeholder confidence. A delayed switch to reduced inspection lowers efficiency as more inspections are done than are needed, while a delayed switch to qualification (or tightened) inspection could result in an increased risk of introduction via leakage.

Timing for paperwork submission and processing. Depending on when paperwork (the manifest, waybill, etc.) for incoming consignments is submitted, processing timing (i.e., determining the inspection status of the lot) can get complicated. If an NPPO begins its process upon arrival of the

consignment for inspection, then its inspection status should be valid, and no mistakes should occur. If, however, the submission of documentation arrives a day or two before the actual consignment, the inspection status might have changed in the interim, and mistakes may occur. Possible mistakes that can occur include (i) skipping a lot that should have been inspected, or (ii) inspecting a lot that might have qualified for clearance without inspection. The first mistake is one that NPPOs will want to avoid. While under-inspections and over-inspections might balance out over time, no mistakes should be tolerated as they may undermine confidence in the program.

The best solution is to determine inspection status only upon arrival of the consignment. Data systems that closely mimic the work-flow process would do this automatically. A less than ideal solution would be to require importers to submit the required paperwork on the day of consignment arrival. However, there may be good reasons (e.g., special targeting operations, staff resource planning) to encourage early submission of paperwork.

System integration. Dynamically determining inspection levels requires integration with the data system that processes incoming consignments and collects information about them (hereafter, "customs system"). However, it is not necessary for the customs system to have real-time access to data on inspection outcomes, nor be integrated with the chosen sampling schemes. For example, the customs system could look-up inspection levels in a separate master table created using results data. Likewise, once the inspection level for a lot is known, inspectors could use a separate tool to determine sample size and selection information. Integration within one system—processing customs data and sampling scheme information—would simplify access for inspectors, but separate systems may be more technologically feasible, especially when beginning an RBS program.

Electronic information. Trade data is quickly moving to electronic transfer of information about goods and consignments. Trade facilitation efforts will streamline the relationship between Customs and other authorities to expedite the movement, release, and clearance of goods in commerce. An example of this is the single window concept, which allow parties involved in trade and transport to provide standardized information and documents at a single entry point to fulfil import, export, and transit-related regulatory requirements (Economic Commission For Europe, 2005). Electronic data should be submitted only once. NPPOs would be well advised to begin developing data systems that can provide large amounts of high-quality data about consignments moving in trade. See United Nations (2011) or WTO (2014) for more information on trade facilitation and the single window.

5.3.2.4. Alternatives to dynamic information systems

Agencies that are unable to develop a dynamic information system, can still implement an RBS program. This may mean limiting the number of pathways/commodities in the RBS program to a number that can be supported with available resources.

In this scenario, the basic approach would be to collect and analyze inspection results at specified intervals (daily, if possible), and then adjust the inspection levels. This approach would mimic some of the simple procedures described in **Table 8**. Many of the tasks could *still* be automated. For example, inspection results could be entered into a spreadsheet template that automatically determines new inspection levels and formats the date for use.

Implementation of the simpler system is at the discretion of the NPPO. If few plant products are eligible, import volumes are low, and ports-of-entry are few, then inspectors might get by using paper copies of current inspection levels. In most cases, it would be preferable to access/manage the information electronically.

5.4. Case studies to illustrate the RBS program design process

- Throughout this chapter we will use three case studies to illustrate the RBS program design process. The first case study country is **Orchard Isles**, a small country with a moderate amount of plant product imports, but a high level of biosecurity concern. The Orchard Isles NPPO (OI-PPO) samples incoming plant products at a high level and is hoping that an RBS program might allow them to further reduce pest leakage.
- The second case study country is **Pasturio**, a moderate-sized country that imports a moderate amount of plant products and relies heavily on exports of animal and dairy products. Their regulatory authority (Pasturio-PO) wishes to optimize inspections especially on the limited plant products they import.
- The third case study country is **Urbania**, a large country with many imports, that wishes to improve safeguarding across a wide array of agricultural products. The Urbanian animal and plant protection organization, Urbania-APPO, currently samples incoming agricultural products at a low rate, and wants to use an RBS program to efficiently increase the intensity of inspections.

5.5. Designing the Risk-Based Sampling (RBS) program

Perhaps the most challenging task for an NPPO is figuring out exactly what a future RBS program will look like. NPPOs may have too many commodities to consider, too many ports-of-entry to focus on, and too many technical issues to align. NPPOs may find it much easier to begin by focusing on a portion of the possible trade pathways and establish a pilot program. This approach will allow NPPOs to develop expertise and experience with RBS designs. As mentioned previously,

building an RBS program is divided into three activity areas - **design, implementation, and maintenance** and the process has 13 steps (**Figure 18**). A 14th and optional step supports major program revisions.

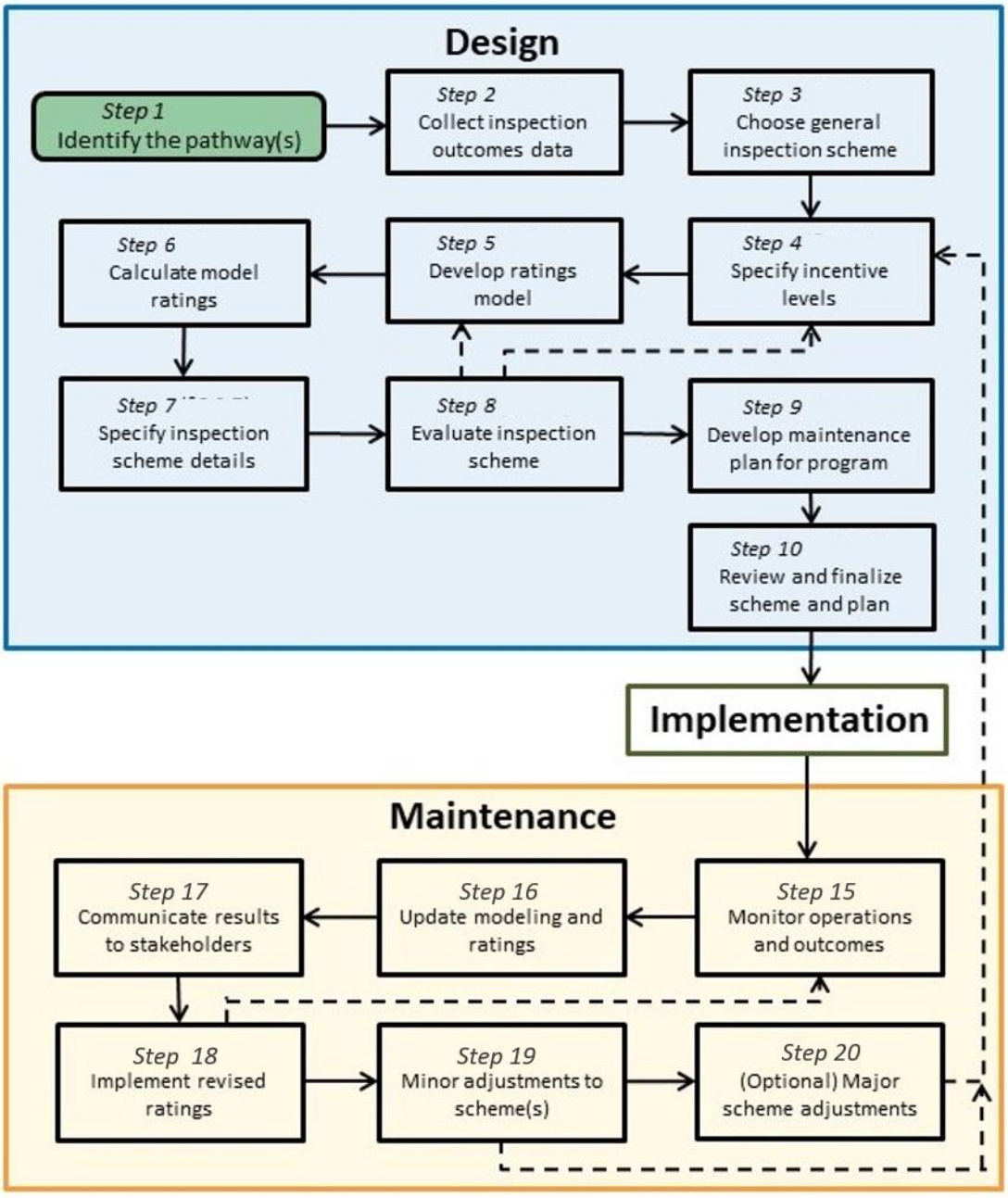


***Figure 18**. Flowchart of the process for **designing, implementing, and maintaining** an RBS inspection scheme. Parentheses list the relevant steps in the text. Dashed lines represent optional paths.*

5.5.1. Step 1 – Identify pathway(s)

5.5.1.1. Considerations

Factors to consider when evaluating trade pathways as possible candidates for an RBS program may include:

- **Trade volume**. An RBS program will be most beneficial when used for high volume commodities, as potential resource savings for inspections is higher.
- **Risk**. It might be best to initially focus on commodities known to be lower risk, to limit the consequences of start-up difficulties or short-term increases in leakage before the quality of the traded commodity improves.
- **Interest from industry**. If producers/shippers are interested in the program then acceptance of new procedures and motivation to improve the quality of the traded commodity will be strong.
- **Resource intensive pathways**. NPPOs may find it useful to focus on pathways that require most of their inspection resources.
- **Other operational aspects**. For example, considering which and how many ports-of-entry will be involved; taking into account involvement of other agencies outside of the NPPO, or the existence of other ongoing special programs at ports-of-entry.

5.5.1.2. Describing and tracking eligible commodities

In addition to selecting the pathway(s), NPPOs must describe the commodities eligible for the RBS program to understand how they will be tracked. Commodities could be identified using taxonomic placement or a tariff code and should include origin.

Specifying commodity identity

- **Category/Type**. If products have well-defined types, these can be used for tracking. As an example, the U.S. NPPO established an RBS program for propagative plant material using 10 different commodity types (e.g., seeds, unrooted cuttings, rooted plants).
- **Taxonomic family**. Using family may be appropriate for commodities that present similar pest risks, such as for flowers (Rosaceae, Liliaceae).
- **Genus**. This may be the simplest identity option. However, using this option might aggregate multiple species that may have different pest risks.
- **Species**. Perhaps the most obvious identity option for tracking as the pest risk can be tied to species. Aggregation may be a problem for specialty crops.
- **Codes**. Harmonized Tariff Codes are used to track and identify commodities in trade. However, they may not be sufficiently detailed (e.g., fruits and nuts) for an RBS program.

The United States uses the harmonized tariff code (Globally Unique Product Identification Code) in addition to scientific name to identify commodities. Identification systems that rely on a combination of taxonomy and product type may be more common. For example, cut flowers are routinely identified by family association (roses, tulips, lilies, etc.), but doing the same for fruits and vegetables (e.g., cucurbits, beans, citrus, berries) might obscure differences in pest risk across species.

It might be challenging for data systems to work with scientific names, especially if importers do not provide that information in a standardized way. Some NPPOs may also make distinctions down to plant variety for certain pest risks, which can complicate tracking.

In some cases, commodity descriptors may negatively impact RBS programs. For example Australia implemented RBS for a dozen or so low risk commodities identified only with tariff codes (Brent, 2016). Tariff codes can be quite specific for some items but quite general for others (e.g., fruits and nuts). The Australian NPPO determined that tariff codes presented challenges when their RBS program was going to be expanded, so they created a new data system that accepted more detailed codes or taxonomic identifiers.

Specifying commodity origin

- **Country**. It is the simplest, but perhaps the least useful commodity origin, because commodities of high and low quality may be grouped together. However, for some pathways, such as sea containers, using country makes sense.
- **Consignee/Broker/Importer**. Receivers of consignments may be entities in the importing country or might also operate in the exporting country. Even though these entities are less able to affect commodity production and quality, they have interest in having reduced inspections. One way for entities to influence the nonconformity rates of consignments is to switch from working with low to higher quality commodity producers. Entities may import a single commodity from different countries, so origin may still need to be tracked.
- **Shipper/Conveyor**. These are entities that transport the commodities. They may be based offshore or operate in the exporting and importing country. They have little control over commodity production or quality but may still be interested in having reduced inspections. If shippers are active in multiple countries, then origin may still need to be tracked.
- **Processor/Exporter**. These are typically local entities, making them well suited for aggregation in order to gain program eligibility and also for influencing commodity production and quality. Note that information on processors/exporters may not be available for RBS program management.

- **Producer/Farm**. This information rarely is available for RBS program management. It is ideal to have in order to provide feedback on commodity non-conformities but may result in too many combinations that cannot qualify for reduced inspection because of their low volumes. Unless producers are directly involved in the export process, they may have little interest in qualifying for reduced inspections.

5.5.1.3. Case study updates

- **Orchard Isles**. The OI-PPO wishes to bring all plant products into one or more RBS programs, perhaps with different sampling schemes for different product types. They plan to begin regulating these items at the level of (country of) origin and genus, but hopefully move to a more detailed level (than origin), such as shipper or consignee in due course.
- **Pasturio**. The Pasturio-PO is focused on inspecting plant products, to provide biosecurity for its own production for domestic use and for exports.
- **Urbania**. The Urbania-APPO is tasked with safeguarding a great number of pathways – far too many to implement an RBS program all at once. As such, they have chosen to begin with cut flowers and all kinds of fruits and vegetables (fresh, frozen, and processed).

5.5.2. Step 2 – Choose a general inspection scheme

In this step, a decision is made in terms of what incentives to offer as part of the RBS program design. Reduced inspection is the typical incentive, but this can be done in one of two ways: reduced **intensity** or reduced **frequency** of sampling.

5.5.2.1. Reduced intensity

Reduced intensity means taking smaller sample sizes. It is the most conservative choice for an RBS program design because every incoming lot is still inspected (that is, *f* stays constant at 100 percent). In certain circumstances, this can result in a faster detection of a change in plant product quality.

Time savings result from inspecting fewer units in lots that are subject to reduced inspection, and fewer total units inspected in a series of lots. For example, in a normal inspection with n = 48 (N = 1,000), if the product is boxes of fruit with 10 pieces per box, then 48 boxes (480 fruit) need inspection. In a reduced inspection under MIL-STD-1916, with n = 20, only 20 boxes (200 fruit) need inspection, a reduction in effort of 280 fruit inspections (58.3 percent). Hence, if inspecting 480 fruit takes one hour, it might only take, 25 minutes to inspect 200 fruits, a savings of 35 minutes. However, each box still has to be 1) held and brought to the inspection area, 2) disassembled for sample selection, 3) inspected, 4) reassembled, and 5) cleared, so time savings may be limited to 35 minutes. Consequently, reduced intensity schemes will usually result in less

time savings than the reduced frequency of sampling approach (see below). This may make it a somewhat less desirable RBS scheme to trading partners.

Another challenge with reduced intensity schemes is their effect on inspection data. Reducing intensity of sampling alters the sampling parameters: if using hypergeometric function parameters, then C or *p*, or both change; when using the OC curve approach, then AOQ or AOQL change. This means that sampling will be variable over time, and the corresponding data will not have the same sample design, which should be taken into account during data analysis.

Furthermore, reducing sampling intensity can lead to very low likelihoods of detecting a nonconformity in a single lot. For example, if only one unit in 1,000 is nonconforming, then n = 48 (see above) results in about a five percent chance of detection ($1 - (1 - 0.001)^{48} = 0.047$). At some point, taking such few samples for relatively large lots will not be worthwhile because detection probabilities are very small.

5.5.2.2. Reduced frequency

In a reduced frequency of sampling scheme, such as the skip-lot standard introduced above, inspection frequency, *f*, starts at 100 percent but decreases under reduced inspection. Sampling parameters remain unchanged, which is advantageous because sampling is consistent (i.e., constant *C* and *p*, or specified values of *AOQ*/*AOQL*), unlike in reduced intensity inspection schemes.

A benefit of reduced frequency schemes is time savings. In this type of scheme inspectors can skip at least four steps (see **Table 9**): disassembly (step 2), sample selection (3), inspection (4), and reassembly (5). The amount of time saved can be significant - more than twice the time savings than in a reduced intensity scheme.

Skip-lot sampling is appropriate when producers/shippers have documented their quality systems and can be trusted (e.g., ANSI/ASQ, 1996), that is, when there is external and verifiable evidence of quality in addition to inspection history. This may not be the case for some pathways, but production practices for a given commodity are generally consistent for similar production locations and growers, and phytosanitary certificates and other documentation that accompany consignments may report processes that maintain product quality standards. However, documentation may not be uniformly reliable. Some NPPOs may have data that indicate that pest action rates are sufficiently low in particular pathways that would favor the implementation of skip-lot sampling. Alternatively, NPPOs could use tables that specify the number of lots to inspect (different from *i*, the clearance interval) before entering a qualification phase for reduced inspection (Table I in ANSI/ASQ, 1996).

***Table 9**. Comparison of general steps and time taken for each under normal and reduced sampling schemes, comparing overall time savings. Times are estimates for the purpose of this example but note that times are the same for both normal schemes. Values in **bold** differ between normal/reduced.*

Inspection step	Reduced Intensity		Reduced Frequency	
	Normal	Reduced	Normal	Reduced
	Time taken (min)			
1. Hold consignment and inspect documents	15	15	15	15
2. Dock and unload consignment	15	15	**15**	**0**
3. Disassemble and select samples	**15**	**10**	**15**	**0**
4. Inspect samples	**60**	**25**	**60**	**0**
5. Reassemble and reload	15	15	**15**	**0**
6. Clear	15	15	15	15
Total time taken	135	95	135	30
Savings under reduced inspection	40		105	

Another scenario to consider is a skip-lot scheme based on CSP-3 (a type of continuous sampling), as described and tested by Robinson *et al.* (2012a) (for CSP-3 also see section 4.5). In this approach, a single non-compliance does not trigger an immediate return to normal inspection levels, but instead mandates tightened short-term inspections, for the next, say, four consecutive lots. Jones *et al.* (2017) refer to this as the alert phase. Only if a second non-compliance is found during the alert phase does inspection revert to normal levels. The number inspections after a non-compliance becomes an additional parameter to specify for the reduced frequency scheme; commonly four are recommended. This approach is likely to minimize inspections compared to a stricter rule switching approach. The approach may be useful for pathways in which producers and importers have occasional non-compliances as opposed to sustained increases in the non-compliance rate.

5.5.2.3. Comparison of plan types

A question that NPPOs might want to consider when comparing reduced intensity versus reduced frequency approaches is - which approach detects problems faster? The answer relies on the exact circumstances of the situation. For example, if the target problem is a ten-fold or more increase in the number of infested units contained in an average lot, then skipping some lots before inspection could result in considerable pest leakage; even at a reduced intensity, the problem might be detected early. On the other hand, if the target problem is less significant, for example a doubling in infestation rates to only 2 percent, then the chance of detecting this change at a reduced intensity is likely to be small (or at least smaller than with normal sampling schemes). In the latter case, we would be more likely to detect the problem at the greater sampling intensity in a reduced frequency approach, even if we skipped some lots before

detecting the positive. It is difficult to favor one approach over the other in this regard. Intuition might suggest that inspecting every lot is an advantage, but it depends on the circumstances of an NPPO. In any case, NPPOs may become more comfortable about the performance of either scheme by conducting a simulation experiment, ideally using existing inspection data (see Robinson et al., 2012a).

Neither approach is simpler to operationalize than the other. For example, in a reduced intensity scheme we know that every incoming lot will be inspected, but the plan to determine the sample size may vary. In a reduced frequency scheme, randomly determining whether or not to inspect skipped lots is an extra step in the process, but the reference sampling plan never changes. Each scheme has its benefits and challenges.

Generally, resource savings or optimization should be greatest with reduced frequency RBS programs (skip-lot sampling), and this is an important factor to consider. If a reduced frequency program is a good fit for an NPPO to implement, it should be done. Greater computational or data system needs can be justified with the expected gains in efficiency. Industry partners seem to prefer them over reduced intensity schemes, primarily because of time savings, but these schemes also reduce handling time and possible harm (or complete loss) during inspection.

5.5.2.4. Case study choices

- The **Orchard Isles** NPPO (OI-PPO) chose skip-lot sampling to maintain high inspection efforts while reducing overall pest leakage.
- **Pasturio** chose MIL-STD-1916 as their standard inspection scheme for incoming plant products, because they believe stakeholders would not favor skipping lots.
- **Urbania** is interested in using either inspection scheme - whichever performs best.

5.5.3. Step 3 – Specify inspection scheme details

Specifying the parameters for a sampling scheme will likely be an iterative exercise and will be refined when the impact of these choices on inspection operations and effort (Step 4) are evaluated. A few different schemes might be assessed before settling on the final specification.

We recommend selecting a standard sampling plan, but NPPOs may also design non-standard plans, or adapt standard plans to their specific needs. The latter could be done by adding or deleting reduction levels or altering parameter values. For example, with ISO 2859-3, an NPPO could alter the four specified *f* values, or add a fifth reduction level, or invoke only one, two or three reduction levels. Likewise, MIL-STD-1916 only specifies reducing inspection by one level, but an NPPO might justify adopting a plan with two or more reduction levels (note that it is not possible to go below verification level R). Below we describe how to design non-standard plans

but more details can be found in these references (Chapter 4 in Stephens, 2001, or Chapter 19 in Schilling and Neubauer, 2017).

When designing the inspection scheme, NPPOs might consider behavioral aspects of program choices (see Rossiter and Hester, 2017; Starbird, 2000; Wan et al., 2013). Stakeholders may respond differently to incentives for increased compliance (quality), depending on the pathway and their flexibility, and some schemes or combinations of parameters may increase their willingness to comply. Not much research has been done on acceptance sampling plans from a regulatory standpoint, and little data exists on how program choices generally affect nonconformity rates for the pathway over time. Surveying stakeholders may serve to gain insight into how to incentivize compliance (see, for example, Rossiter and Hester, 2017).

Keep in mind that any inspection scheme to be implemented by an NPPO needs to be easy to explain, not only for those responsible for implementing it (the inspectors) but also for the stakeholders that will benefit from its uptake.

"If you want a method or system used, keep it simple." (Dodge, 1977)

5.5.3.1. Inspection reduction levels (incentives)

The first step is to determine how many different levels of reduction (or tightening) to use in the program. This is the same as determining how many levels of incentives (or disincentives) will be offered to importers and shippers. Having fewer levels is preferable, both for communication to importers and for ease of data system programming.

For example, the ISO standard 2859-3 specifies four levels of skipping inspection (ISO, 2005) while the standard MIL-STD-1916, allows for only one level of reduced inspection below the prescribed level (Department of Defense, 1996); the reduced sample size is about 60 percent less, regardless of the default verification level. Other plans will require further specification. For example, the SkSP-2 inspection scheme has no specified levels - users themselves define the number of reduced inspection levels and associated parameters.

5.5.3.2. Sampling scheme parameters

A. Reduced intensity schemes

Standard plan. If using MIL-STD-1916, or a similar reduced intensity scheme (se Z1.4; ANSI/ASQ, 1993), then $i = 10$, and the only choice remaining is the normal (baseline) verification level (VII-I). That information determines the reference sampling plan for the inspection scheme. The choice is based upon the desired quality level, most likely using *AOQ* or *AOQL* values. The curves can be viewed (by verification level) in the Appendix of the Handbook for MIL-STD-1916 (Department of

Defense, 1999); below we show one example for verification level III, with *AOQL* values over five lot size codes - from about 0.45 percent to 1.1 percent (**Figure 19**).

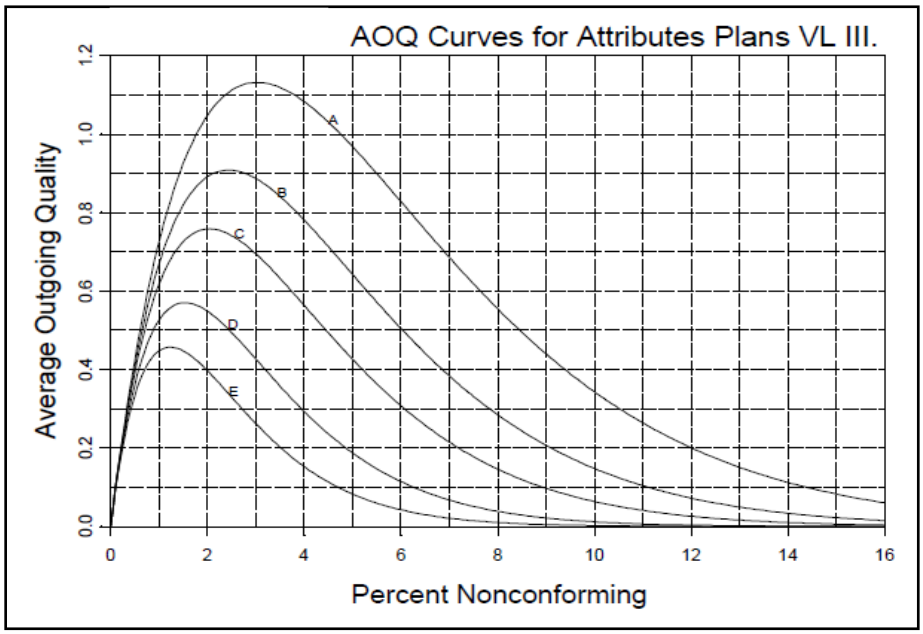


***Figure 19**. AOQ (percent) curves by lot size (A-E) versus percent nonconforming lots for the MIL-STD-1916 plan at verification level III. Values for AOQL can be read from the maximum value of each curve or can be calculated directly (see Appendix A). Figure attribution Department of Defense, 1999.*

In hypergeometric terms, if one specifies C = 0.95 and works backward from the given sample sizes, then verification level III equates to p = 0.089 at lot size A and to p = 0.037 at lot size E (see **Table 4**). For comparison, the sample sizes given by verification level II are p = 0.22 at lot size A, to p = 0.089 at lot size E.

The MIL-STD-1916 handbook recommends the use of verification level VII for risks deemed critical, whereas levels III to VI are recommended for major risks. Minor risks can be managed using verification levels I to III.

Non-standard plan. If an NPPO decides to add one or more reduced inspection levels to a standard plan, they should specify the normal verification level (reference sampling plan) and confirm that a constant i is being used for all switching rules. If varying values for i for different reduction levels are chosen, then those need to be justified. For example, an NPPO might use MIL-STD-1916 and specify a normal verification level of IV, with i_1 = 15 for switching to level III, but also specify i_2 = 20 for a further reduction to level II. The justification might be that they wish to ensure greater quality before effecting the second switch.

If the reduced intensity scheme is being built de novo, then the NPPO needs to specify:

- *i*, which will either be constant across all reduction levels (recommended) or specified uniquely and justified for each level; and
- when using *C* and *p* (hypergeometric functions):
    - Specify *C* and *p* for the normal inspection level
    - Specify *C* and *p* values for each reduction level
- when using *AOQL*/*AOQ* (OC curve approach):
    - Specify desired value of *AOQL* (or alternatively, the mean *AOQ*) for the normal inspection level[13]
    - Specify desired values of *AOQL* (or mean *AOQ*) for each reduction level.

Note that whichever approach is used, the parameter values for the normal inspection level define the reference sampling plan for the inspection scheme.

B. Reduced frequency schemes

Standard plan. The first step in specifying any skip-lot plan is to define the reference sampling plan. This can be done using either the OC curve or the hypergeometric approach. A standard plan will often include set values for *i* and *f*. The value of *i* may be constant, but there are usually multiple *f* values that specify the steps for skipping inspection levels. For example, if using ISO 2859-3, the parameters are set at *i* = 10, and *f* = 0.5, 0.4, 0.3, and 0.2. If using SkSP-2 or another skip-lot scheme with no proscribed parameters, NPPOs should follow the guidance for non-standard plans.

Non-standard plan. Here as well, the first step is to define the reference sampling plan using the OC curve or hypergeometric approach (or both). The next step is to evaluate how different *i* and *f* combinations affect the detection level of the reference plan and find one that meets NPPO expectations for time to qualification (*i*) and fraction inspected (*f*) and gives the desired *AOQL* (or mean *AOQ*).[14]

Note that lower values of *i* can be compensated for by greater *f* values and vice versa, so there are many combinations that could meet the desired specifications of the NPPO. The consideration for choosing the value of i is how long the NPPO can wait until some products begin qualifying for skipping inspection, versus how much assurance they want to provide to stakeholders that quality is sufficiently high to allow skipping inspection. Similarly, some considerations for choosing *f* are how much the NPPO needs to reduce the overall inspection effort (resource savings) versus how comfortable they are with allowing many lots to be accepted before a significant change in quality is detected.

[13] Stephens (2001) provides an equation for calculating n based on AOQL. See Appendix A.
[14] Or relate to the parameters *C* and *r* if preferred.

Quantifying the effects of different values of *i* and *f* on the nonconformity rates achieved in conjunction with the reference sampling plan is not a trivial task, especially for *AOQ* because it varies with both *p* and *n*. The acceptance sampling reference books by Stephens (2001) and Schilling and Neubauer (2017) provide spreadsheets useful for this purpose. Reference tables for $AOQL_2$ are also available in Perry (1973) [the relevant table is reproduced in both reference texts]. Such tables are limited to the listed *i* and *f* values. Simulation exercises based on historical data or models can also be used based on the expected or observed patterns of pathway contamination (Robinson et al., 2012a).

Note that skip-lot sampling decreases the quality level provided by a reference sampling plan, because some nonconforming lots are being accepted without inspection. However, the savings in inspection effort and resources can be redirected to manage other risks.

5.5.3.3. Sample size determination by inspectors

Regardless of which inspection scheme is used, NPPOs must specify how sample sizes will be determined by officers/personnel carrying out inspections. Options include using tables (existing or newly developed), manual calculations, or using a special calculator tool or spreadsheet (see Appendix for functions). These can be made available in print or in digital or online formats to inspectors.

For example, when using MIL-STD-1916 without modification, providing paper copies of Tables I (lot size codes) and II (sample sizes by code and verification level) may be enough for inspectors to determine the sample size. For other schemes, tables to determine sample size may have to be developed by the NPPO specifically for their chosen inspection scheme.

For reduced frequency plans, the reference sampling plan will rarely change. It can be provided in the format best suited for the operational situation (as above). However, the NPPO may need to provide a tool to randomly determine whether to inspect a lot that is in the skipping phase. See Annex B in ISO (2005) for suggestions on how to do this using random number tables or dice.

5.5.3.4. Case study choices

Orchard Isles. OI-PPO decided to use skip-lot sampling. The reference sampling plan was set to achieve very high-quality, based on an AOQL of about 0.00045, which equates to $C_{Ref} = 0.95$ and $p_{Ref} = 0.005$. Under this plan, a lot size of 1,000 would have a hypergeometric sample size of 450.

OI-PPO decided to adopt ISO 2859-3 for most product pathways. As such, the parameters used were as listed in the standard plan (see 5.2.2.2). The estimated quality levels achieved with this plan are $AOQL_1 = 0.00082$ (regardless of f), while $AOQL_2$ varies from 0.00262 at f = 0.5 to 0.00661 at f = 0.2. Note that levels are very low because of the high intensity of the reference sampling plan.

For low-risk pathways, such as that for processed products (e.g., dried fruit) or items for processing (e.g., coffee beans), OI-PPO used SkSP-2. Because the pathways were low-risk, OI-PPO felt comfortable specifying f values that stepped down from 0.1 (1-in-10), to 0.05 (1-in-20), and finally as low as 0.02 (1-in-50), always with i = 10. The value for $AOQL_1$ was unchanged from above, but $AOQL_2$ increased to 0.0995 at f = 0.1, and up to 0.184 at f = 0.02.

Pasturio. Pasturio PO decided to use MIL-STD-1916 with verification level IV for their normal sampling plan and two levels of reduced inspection after i = 12. The mean sample size for level IV is 131 (range = 80 to 192) (Department of Defense, 1996), corresponding to a value for p_{Ref} of about 0.021 when C_{Ref} = 0.95 (N = 1,000), $AOQL_1$ = 0.00049, and $AOQL_2$ = 0.0024.

Urbania. For most pathways Urbania inspects at low intensity. They wanted to at least double the intensity on most pathways, with inspections capped at the current level of effort. Their current sampling is set at p_{Ref} = 0.15 and C_{Ref} = 0.95, which gives n = 19 for N = 1,000, and equates to $AOQL_1$ = 0.0150 and $AOQL_2$ = 0.0190. To double the sample size to 38, one can decrease p_{Ref} to 0.075, which equates to $AOQL_1$ = 0.0057 and $AOQL_2$ = 0.0093.

Urbania decided to compare MIL-STD-1916 against the SkSP-2 plan for cut flowers. For MIL-STD-1916 they used verification level II as normal, because the mean n at that level was about 21 and the mean AOQL about 0.034, over all lot sizes (after Schilling and Neubauer, 2017). For SkSP-2, they chose f = 0.6, to align with the 40 percent reduction in sample size and used n = 21 instead of the baseline value of 19.

Based on these results (see below), they specified two different SkSP-2 plans for fruits and vegetables, one for fresh produce, and one for frozen/processed items. The plan for frozen/processed items had i = 10, with f values of 0.5, 0.3, 0.1, and 0.05. This results in quality levels of $AOQL_1$ = 0.0189 (all f), and $AOQL_2$ = 0.0262 (0.5), 0.0477 (0.3), 0.0995 (0.1), and 0.135 (0.05). Because of the low risk associated with frozen/processed items, they opted not to increase sampling intensity for this pathway. The fresh produce plan was similar (f values of 0.5, 0.25, and 0.125) and was subjected to increased sampling intensity.

5.5.4. Step 4 – Evaluate the chosen inspection scheme

Evaluating the impacts of the inspection scheme and its specifications on the overall process is important. Evaluation should include the level of effort required, its effect on the number of non-conforming lots detected, and the corresponding leakage/slippage. Pre-implementation evaluation is an excellent way to assess some expected outcomes and ensure that the goals of the program are met.

It might be reasonable to evaluate more than one inspection scheme (Step 3, above) on its processes, operations, and outcomes. The process of finding a scheme that meets the desired

outcomes for an NPPO could be iterative. Each NPPO should determine what values must be met for the inspection scheme to be considered acceptable. In general, effort levels should meet or exceed expectations and efficiently use resources (time taken for inspections) while minimizing leakage rates.

5.5.4.1. Calculate performance indicators

Below we present ways of estimating total inspections and samples taken, time required, and leakage for the two different types of inspection schemes. Other metrics may be of interest and could be developed by NPPOs themselves. Arthur *et al.* (2013) has additional suggestions.

A. Reduced intensity schemes

Inspected samples. The most direct method to evaluate how the inspection scheme affects the effort (or resources) required is to estimate the total number of samples taken during normal and reduced (and tightened, if appropriate) inspections (Appendix D1) and compare these data to the number of samples taken when using a non-RBS approach.

Instead of using the mean *L* over all lots, they could be separated into cohorts of similar size, such as lots with mean sizes of 500-1,000, 1,000-2,000, etc., or different levels of *p* or *d*. After calculating the results for each cohort, sum the results to find totals. This is more complicated but may improve the estimates.

Time taken for inspections. Here we estimate and compare how long (person-hours) it takes to perform inspections under RBS and non-RBS approaches. Consider how long each step in the entire process (arrival to clearance) takes per consignment, and then determine the total time over all consignments under RBS and non-RBS approaches (Appendix D2). A next step might be to calculate how many persons would be needed to achieve these totals under a normal work schedule.

Leakage. The objective is to compare the number of defective (infested) units that escape detection under RBS and non-RBS approaches.[15] The value will usually be greater under RBS because fewer total lots, or fewer total samples, are being inspected. If, however, the RBS program being evaluated increases *n* for higher risk items, perhaps capping the overall inspection effort, then leakage might be similar or perhaps decrease under RBS and non-RBS approaches. The approach to evaluate leakage is to estimate the total number of non-conforming (infested) lots that are accepted and use those values to estimate the number of defective (infested) units accepted under each scenario (Appendix D3).

[15] This concept is different than the idea of a 'leakage survey,' which means post-inspection checks to determine the effectiveness of just-completed inspections (see Robinson *et al.*, 2012b).

In this approach we have estimated the number of defective *items*, or those infested by pests, rather than the number of pests or propagules themselves. Pest propagule pressure is, however, the most important biosecurity risk metric (Blackburn et al., 2011; Williamson and Fitter, 1996). One could estimate pest propagule pressure from the mean number of propagules per infested item, if known. However, this number varies by pest type, by commodity type, just to name a few important factors. As such, estimating the number of pests or propagules accepted is more difficult and uncertain than estimating the number of defective items accepted.

B. Reduced frequency schemes.

For many skip-lot sampling schemes, possible status will be either normal or reduced inspection, so the methodology should determine the proportion of lots under each status. When using a skip-lot scheme based on CSP-3 (see Robinson *et al.*, 2012a) (for CSP-3 also see section 4.5), the proportion of time spent in the alert phase, with normal inspections, also needs to be estimated.

Inspected samples. The method differs from the approach described above in that sample sizes do not change, but we must account for the fraction of lots that are not being inspected (Appendix E1).

Time taken for inspections. The method is like the one described above, but the potential time savings is greater because when lots are cleared without inspection some steps are skipped altogether (**Table 9**) (Appendix E2).

Leakage. As above, we need to estimate the number of non-conforming lots that are accepted under RBS and non-RBS approaches, and then estimate and compare the number of defective units in those lots (Appendix E3).

5.5.4.2. Specialized analyses

Person-hours/Staff required. If a direct assessment of resources in person-hours is needed, then metrics should be converted to time-based units. For example, the number of samples taken can be converted to person-hours using an estimate of mean time taken to inspect each sample. The time required for the entire inspection process can be estimated by adding the time required for activities other than sampling (e.g., off-loading the material, selecting the samples).

Using saved resources for more intensive inspections. Reinvestment of saved resources to increase inspection intensity is also possible. The approach is to estimate the number of inspections (or time required) under the current (non-RBS) system, then apply the RBS approach to estimate the savings gained, and then increase sample sizes (or frequencies) until the current effort levels are approximately matched. Such reinvestment is not always done.

Simulation-based approaches. Simulations are more resource-intensive but may provide more realistic estimates than the approaches used above (see Springborn et al., 2018). This is because

simulation-based analyses can explicitly account for uncertainties. Also, because inspection outcomes are inherently probabilistic, a simulation approach suits the process very well.

5.5.4.3. Case study examples

Orchard Isles. OI-PPO chose to use skip-lot sampling on all pathways.

Very low-risk pathways. OI-PPO identified 25 commodities typically arriving in lots of about 5,000 units. The reference sampling plan for this lot size gives n = 564. They estimated they could expect 50,000 lots annually (2,000 per commodity on average), and estimated p = 0.1 (non-conforming fraction) and d = 0.0002 (fraction defective) (**Table 10**, below).

Almost 750,000 samples were taken annually under the RBS approach (**Table 11**, below). This is a 97 percent reduction when compared to the non-RBS approach, while leakage only increased by an estimated 54 units. Also, the time required for all operations under RBS was just over 26,000 person-hours, compared to 65,000 person-hours under the non-RBS approach, a decrease of 60 percent.

Other pathways. OI-PPO used the ISO 2859-3 plan by dividing the pathways into groups with similar lot sizes. Fresh produce (75 commodities) had L = 1,000 per commodity, mean N = 1,500, and mean d = 0.001 (**Table 10**). Cut flowers and some other products (60 total) had N = 5,000 and total L = 2400. Propagative material (25 total) had N = 10,000, with total L = 1,200.

For fresh produce, about 30M samples were taken under RBS, compared with about 37M under the non-RBS approach, a savings of 19 percent (**Table 11**). The time taken for inspections dropped by 15 percent under the RBS approach. Leakage increased under RBS by 1,194 units, or a proportional increase of 3 over the total without RBS. That meant that almost 5,900 samples were saved per unit increase in leakage. The relatively low savings from RBS in this category reflects that relatively large n in combination with relatively small N (which lowers P_a) tends to limit the potential of commodities to get to lower f levels.

For cut flowers and other products the total number of samples taken under RBS was just over 500K, compared to over 1.35M samples without RBS, a savings of about 47 percent (**Table 12**). The time required for inspections without RBS was 4,800 hours but dropped by 47 percent to 2,534 with RBS. Leakage increased by only 7 units, or 1.5 percent, which saved more than 121K samples per unit increase in leakage.

Lastly, for propagative material total samples taken under RBS were about 255K, compared to almost 700K without RBS, a decrease of 63 percent (**Table 12**). Person-hours dropped from 2,400 to about 1,250 under RBS, a 48 percent decrease. Leakage was estimated to be zero, reflecting both the low volume and the low non-conformity rate on the pathway.

Combining the results from all pathways, the Orchard Isles PPO took 35.7M fewer samples to clear all the lots, a savings of 53 percent. This saved over 67,000 total person hours, at an increase in leakage of only 1,255 units, or over 28,500 samples saved per unit increase in leakage.

***Table 10**. Parameter values for evaluating the proposed inspection scheme designs for pathways in Orchard Isles, by product type: processed items and items destined for processing, fresh produce, cut flowers, and propagative materials (PM). For all pathways, AOQL was about 0.00045, which equates to C_{Ref} = 0.95 and p_{Ref} = 0.005.*

Parameter	Description	Values			
		Processed	Fresh	Cut flowers	PM
L	Mean annual arriving lots (no.)	50,000	75,000	2,400	1,200
N	Mean lot size (no.)	5,000	1,500	5,000	10,000
p	Fraction of non-conforming lots	0.1	0.2	0.2	0.1
n_{Norm}	Sample size, normal inspections (no.)	564	493	564	581
d	Mean fraction of defective units per lot	0.0002	0.002	0.0002	0.0001
D	Mean defective units per lot (no.)	1	3	1	1
P_a	Mean P_a, adjusted for p	0.989	0.861	0.977	0.994
P_r	Mean P_r, adjusted for p	0.011	0.140	0.023	0.006
h_{insp}	Person-hours per inspected lot	1.3	2.25	2.0	2.0
h_{non}	Person-hours per non-inspected lot	0.5	0.5	0.5	0.5
i	Clearance number[a]	10	10	10	10

[a] For the number and values of f, see 5.5.3.4.

***Table 11**. Estimated metrics for proposed inspection scheme designs for processed/processing items, or cut flowers in Orchard Isles, with and without RBS. Reported values are estimated mean annual responses.*

Description	Values			
	Processed		Fresh produce	
	No RBS	RBS	No RBS	RBS
Comparative estimates				
Total samples taken for all lots (no.)	28,200,000	749,556	36,975,000	29,932,495
Total time taken for inspections (person hours)	65,000	26,063	168,750	143,751
Total accepted non-conforming lots (no.)	4,944	4,998	12,907	13,305
Total defective units accepted (no.)	4,944	4,998	38,721	39,915
Summary of RBS effects				
Total samples saved using RBS (no.)	27,450,444		7,042,505	
Proportional savings in samples inspected	0.97		0.19	
Total time saved using RBS (person-hours)	38,937		24,999	
Proportional time savings for inspections	0.40		0.15	
Leakage (increase in defective units, no.)	54		1,194	
Proportional increase in leakage of units	0.01		0.03	
Samples saved per unit increase in leakage	508,342		5,898	

***Table 12**. Estimated metrics for proposed inspection schemes for cut flowers and propagative material in Orchard Isles, with and without RBS. Reported values are estimated mean annual responses.*

Description	Values			
	Cut flowers		Propagative material	
	No RBS	RBS	No RBS	RBS
Comparative estimates				
Total samples taken for all lots (no.)	1,353,600	501,396	697,200	255,059
Total time taken for inspections (person-hours)	4,800	2,534	2,400	1,259
Total accepted non-conforming lots (no.)	469	476	119	119
Total defective units accepted (no.)	469	476	119	119
Summary of RBS effects				
Total samples saved using RBS (no.)	852,204		442,141	
Proportional savings in samples inspected	0.63		0.63	
Total time saved using RBS (person-hours)	2,267		1,142	
Proportional time savings for inspections	0.47		0.48	
Leakage (increase in defective units, no.)	7		0	
Proportional increase in leakage of units	0.015		0.0	
Samples saved per unit increase in leakage	121,743		N/A	

Pasturio. The Pasturio PPO assumed that every lot was nonconforming (p = 1.0), and that d = 0.0015. Mean N was 2,000, and on an annual basis they expected to receive about 7,500 lots (about 20 per day, on average) (**Table 13**). Without RBS, they expected to take 960,000 samples (= 7,500 × 128) to clear all incoming lots. The results indicate that by adopting the proposed RBS program the number of samples taken would decrease by 13.5 percent (**Table 14**). Leakage only increased by 528 units, or 2.9 percent over normal expectations (**Table 15**). This represents a savings of about 245 samples per unit increase in leakage, which was judged as acceptable. Pasturio decided to implement the proposed RBS program.

***Table 13**. Parameter values for evaluating proposed inspection scheme designs for pathways in Pasturio.*

Parameter	Description	Value
L	Mean annual arriving lots (no.)	7,500
N	Mean lot size (no.)	2,000
d	Mean fraction of defective units per lot	0.0015
n_{Norm}	Sample size, qualifying inspections (no.)	128[a]
n_{Red}	Sample size, reduced inspections (no.)	48[b]
$P_{a\text{-}Q}$	Mean P_a for qualifying inspections (proportion)	0.82
$P_{a\text{-}Red}$	Mean P_a for reduced inspections (proportion)	0.93
p	Fraction of non-conforming lots	1.0
D	Mean defective units per lot (no.)	3
i	Clearance number	12
h_{insp}	Person hours per inspected lot	1.3
h_{non}	Person hours per non-inspected lot	0.5

[a] MIL-STD-1916, VL = IV
[b] MIL-STD-1916, VL = III

***Table 14**. Evaluation of inspection effort, or number of samples taken, for Pasturio pathways.*

Parameter	Description	Values
F_{Red}	Fraction of lots under reduced inspections	0.215
L_{Red}	Lots in qualifying inspections (no.)	1,615
L_Q	Lots in qualifying inspections (no.)	5,885
M_Q	Total samples in qualifying inspections (no.)	753,280
M_{Red}	Total samples in reduced inspections (no.)	77,520
M_{tot}	Total samples in all RBS inspections (no.)	830,800
M_{Saved}	Total samples saved by using RBS (no.)	129,200
	Proportional savings in effort	0.135

***Table 15**. Evaluation of leakage with RBS, or potential increase in defective units accepted in the Pasturio case study pathway.*

Description	Values
Mean accepted non-conforming lots, qualifying (no.)	4,825
Mean accepted non-conforming lots, reduced (no.)	1,501
Mean accepted lots, total (no.)	6,326
Total defective units accepted (no.)	18,978
Leakage (increase in defective units, no.)	528
Proportional increase	0.029

Urbania. Urbania PPO inspects at low intensity and wants to at least double sampling intensity in most pathways, but cap inspections at current effort levels. Consequently, the approach used two steps: 1) test for RBS with the baseline sampling level, and 2) then test for RBS with double sampling level, to verify that total inspections did not exceed levels under the non-RBS approach.

Cut flowers. The PPO compared MIL-STD-1916 with a skip-lot sampling plan. They estimated that mean N for cut flowers was 6,000, which gives mean n = 32 for the MIL-STD-1916 plan, and n = 38, for the skip-lot plan using hypergeometric sampling. They expect about 25,000 lots annually. In addition, a mean of 20 percent of lots were non-conforming (p = 0.2), d = 0.0005, and inspection time per sample was 0.5 min.

The strictest verification level that could be used without significantly exceeding the cap on samples inspected (i.e., no more than 5 percent) was VL III. At that intensity, the estimate for total samples taken was over 840,000 (**Table 16**). At that sampling level under a non-RBS approach, 2M samples would have to be taken to clear all lots, indicating significant savings under the RBS approach. Moreover, because of the increase in intensity, leakage increased by less than 1 percent in either scheme (**Table 17**). The sample savings per unit increase in leakage was over 386,000. For a relatively clean pathway such as this, using RBS helps maximize sampling intensity and improve safeguarding. A comparison of results for MIL-STD-1916 and SkSP-2 plans resulted

in the Urbania PPO deciding that the sampling and time savings from SkSP-2 more than made up for the increased leakage.

***Table 16**. Evaluation of inspection effort, or number of samples taken, for the Urbania cut flower pathway comparing MIL-STD-1916 and SkSP-2.*

Parameter	Description	Values	
		MIL-STD-1916	SkSP-2
n_{Norm}	Mean sample size, qualifying inspections (no.)	80[a]	38
n_{Red}	Mean sample size, reduced inspections (no.)	32[b]	38
p	Fraction of nonconforming lots	0.2	0.2
$P_{a\text{-}Q}$	Mean P_a for qualifying inspections (proportion)[c]	0.992	0.996
$P_{a\text{-}Red}$	Mean P_a for reduced inspections (proportion)	0.997	0.996
D	Mean defectives per lot (no.)	3	3
F_{Red}	Fraction of lots under reduced inspection	0.966	0.965
L_Q	Lots under qualifying inspection (no.)	847	1001
L_{Red}	Lots under reduced inspection (no.)	24,153	24,115
M_Q	Total samples taken in qualifying inspections (no.)	67,760	38,038
M_{Red}	Total samples taken in reduced inspections (no.)	772,896	38,038
M_{tot}	Total samples in all RBS inspections (no.)	840,656	76,076
$M_{tot,NoRBS}$	Total samples taken without RBS (no.)	2,000,000	950,000
M_{Saved}	Total samples saved by using RBS (no.)	1,159,344	873,924
	Proportional savings in samples taken	0.58	0.92

[a] MIL-STD-1916, VL = III
[b] MIL-STD-1916, VL = II
[c] P_a values have been adjusted for p

***Table 17**. Evaluation of leakage and time taken, comparing programs that reduce inspections by decreasing sample size (MIL-STD-1916) or by decreasing frequency of inspections (SkSP-2) for the Urbania cut flower pathway.*

Description	Values	
	MIL-STD-1916	SkSP-2
Mean accepted non-conforming lots, qualifying (no.)	168	199
Mean accepted non-conforming lots, reduced (no.)[c]	4,815	4,822
Mean accepted nonconforming lots, total (no.)	4,983	5,021
Total defective units accepted (no.)	14,949	15,063
Leakage (increase in defective units, no.)[d]	66	120
Proportional increase	0.004	0.008
Time taken for qualifying inspections (person-hours)	1,626	1,301
Time taken for reduced inspections (person-hours)	29,467	12,858
Total time taken with RBS (person-hours)	31,093	14,160
Proportional decrease from maximum time without RBS	0.352	0.71

Fruits and vegetables. The results of evaluating SkSP-2 for fresh fruits and vegetables are shown in **Tables 18 and 19**. Because the scheme for frozen and processed items was qualitatively very similar, we do not present those results. The results demonstrate what can be accomplished if savings from reduction in inspections achieved through an RBS approach are used to intensify

sampling on the pathway rather than just reducing effort. If sampling occurs at the normal level, the total samples taken could decrease by about 86 percent (**Table 18**), while the time required would decrease by two-thirds (**Table 19**). Instead, sample size is doubled (**Table 18**), which means fewer lots qualify for reduced inspection, but the total number of samples taken is still 50 percent lower than without an RBS approach, and leakage becomes negligible (**Table 19**).

***Table 18**. Evaluation of inspection effort, or number of samples taken, for the Urbania fresh produce pathway using skip-lot sampling.*

Parameter	Description	Values with RBS	
		Standard	Higher Intensity
n_{Norm}	Sample size, qualifying inspections (no.)[a]	19	38
P_a	Mean P_a (proportion)[b]	0.981	0.963
D	Mean defectives per lot (no.)	5	5
F_{Red}	Fraction of lots under reduced inspection (all f)	0.97	0.80
L_{Red}	Lots with reduced inspection (no.)	96,712	80,427
L_Q	Lots with qualifying inspection (no.)	3,288	19,604
M_Q	Total samples taken in qualifying inspections (no.)	62,472	744,952
M_{Red}	Total samples taken in reduced inspections (no.)	208,145	1,146,080
M_{tot}	Total samples in all RBS inspections (no.)	270,617	1,891,032
M_{Saved}	Total samples saved by using RBS (no.)[c]	1,629,383	1,908,968
	Proportional savings in samples taken	0.858	0.502

[a] Hypergeometric
[b] P_a values have been adjusted for p = 0.5
[c] Compared to inspections without RBS at the same n_{Norm}

***Table 19**. Evaluation of leakage, or number of defective units accepted, and total time taken, with skip-lot sampling at normal or enhanced sampling intensity, for the Urbania fresh produce pathway.*

Description	Values under RBS	
	Standard[a]	Higher Intensity[b]
Mean accepted nonconforming lots, qualifying (no.)	1,613	9,440
Mean accepted nonconforming lots, reduced (no.)[c]	48,254	39,658
Mean accepted nonconforming lots, total (no.)	49,867	49,098
Total defective units accepted (no.)	249,335	245,490
Leakage (increase in defective units, no.)[d]	4,015	170
Proportional increase	0.016	0.001
Inspection time (person hours)	32,047	111,969
Clearance time (person hours)	42,879	25,134
Total time taken (person hours)	74,925	137,103
Time savings (person hours)[e]	150,075	87,898
Proportional decrease	0.667	0.391

[a] n_{Norm} = 19
[b] n_{Norm} = 38
[c] Includes lots cleared without inspection
[d] Relative to mean accepted nonconforming lots without RBS (= 245,320)
[e] Relative to the time taken without RBS (= 225,000 person hours)

End of case studies. The case study examples end here, because planning, review, and monitoring are more dependent on particular circumstances (e.g., organizational structure) and on NPPO resources and processes.

5.5.5. Step 5 – Develop a plan to provide feedback on inspection outcomes

In a recent study on the economics of RBS, Australian researchers found that providing feedback to importing entities about inspection outcomes was critical to influence improvements in product nonconformity rates (Rossiter et al., 2016). Unless importing entities are aware of their history of biosecurity compliance, they may not know about their non-conformities, and therefore may not appreciate that they are missing out on available incentives if they take corrective actions.

Importing entities that closely monitor how many non-conforming lots get returned, destroyed, or treated may be more informed, but many of them build such costs and losses into their business plans and may not appreciate the potential savings in time and money. Even those that monitor costs and losses may be unclear on how much their profits or customer satisfaction might improve. Finally, quality improvement may involve costs or disruption to established production practices, so some importing entities may not want to initiate changes unless the benefits are well understood. Therefore, developing an effective plan for communicating inspection outcomes to importing entities is an essential step for ensuring the success of implementing an RBS program, and for maximizing the program's potential to reduce pest risks. Creating the communications plan before implementing RBS is important, so that required information and analytical results can be identified and built in, and processes worked out, before they are needed.

The basic objectives of the communication plan are to 1) summarize recent inspection outcomes, and, if appropriate, 2) meaningfully convey the costs incurred (or the incentives missed) by not being more in compliance. This latter point may be as simple as indicating how many lots received normal inspection instead of reduced inspection. However, creating an effective communications plan involves some technical and procedural issues, and perhaps legal or regulatory issues as well. Some of these may include the following:

- Correctly identifying the importing entities with which to communicate
- Following NPPO guidelines for external communications
- Determining the frequency of feedback
- Summarizing/presenting data in a meaningful way, and at the appropriate level of detail
- Estimating what savings could have been achieved if quality/compliance had been higher
- Automating that process so far as possible

The most important issue is that communications are delivered to an importing entity that is capable of effecting product quality. Leveraging existing business communication processes between importers, brokers, and suppliers may be viable (Rossiter et al., 2016). Third parties may be utilized at times, but direct communications by the NPPO are recommended. The communication/feedback process needs to account for any legal or regulatory restrictions and maintain appropriate levels of security and privacy. NPPOs may have specific units for external communications, so those groups would need to be consulted and integrated into the process.

The NPPO should regularly monitor inspection outcomes (see 5.5.6), and feed information to the communication/feedback effort. The timing and regularity of reports can be tailored to the monitoring process, although long delays between reports limit the ability of importing entities to improve product quality and take advantage of incentives (see section below for more information).

NPPOs need to decide how much information to provide. Is it enough to indicate the number of conforming/non-conforming lots by commodity, or is there interest or need to provide more detailed information about the non-conformities (e.g., pest identity)?

Information automation is critical, because constantly providing information to several importing entities will be challenging. The Australian PPO created a script for the open-source statistical program R that takes relevant results from a dataset, and compiles individualized importing entity reports for mass emailing, including all standard explanatory text (Brent, 2016). One means of managing the scope might be to limit reporting to those importing entities that have met a minimum volume threshold over the last few reporting cycles.

Providing some estimate of the gains the importing entity could achieve if product quality is improved may be important. This could motivate them to comply, especially if their business plans already account for potential losses and treatment costs, and those are not being exceeded. Even if the estimated savings seem negligible—perhaps because most incoming goods on the pathway are already high quality—the feedback could emphasize the importance of maintaining that quality level.

5.5.6. Step 6 – Review and finalize the scheme(s) and feedback plan

This step involves an internal review of the new inspection scheme(s) and feedback plan, and perhaps an external review by or a consultation with stakeholders. This step might be applied iteratively with Steps 2-4. The goals of the internal review are to ensure that the plan is feasible, contains all the relevant information, and is likely to achieve its objective. The review might also educate the NPPO about the program and facilitate the program buy-in process. The latter may be especially important if RBS is being implemented for the first time, because new approaches may be unfamiliar and viewed with some skepticism. Good practices during an internal review

include ensuring the participation of relevant experts and affected NPPO units, and responding explicitly and transparently to all suggestions, comments, and questions.

The goals of an external review would include ensuring that the program and incentives are understood, and that the types and levels of incentives are acceptable and meaningful to stakeholders. Relevant stakeholders might include producers or importers; industry associations (at any level); and partnering agencies. An external review by these stakeholders will allow them to begin planning for any changes in their normal production or shipping practices that may need be altered to utilize the incentives.

5.6. Implement inspection scheme

5.6.1. Step 7 – Rollout planning

The rollout of new RBS scheme(s) could include regulatory updates, creating or updating manuals and other documentation, final programming and tuning of data systems or sample size calculators, and making NPPO organizational adjustments to staffing or assigned responsibilities. In addition, some training and outreach to staff, managers, and stakeholders and other relevant parties will also be needed. It is critical to identify any tasks that need completion before the RBS scheme can be implemented. Having a thorough and complete listing of the tasks will facilitate tracking and completion. Essential tasks and activities might include the following:

- **What** materials need to be developed, updated, or prepared?
  - Regulations
  - Sampling tools or calculators
  - Data system programming requirements or plans
  - Briefing papers/reports justifying/explaining the RBS scheme and its specifications
  - Training and outreach materials
- **What** actions are required before the RBS scheme can be initiated?
  - Approval of the program and commitment from the NPPO
  - Union/employee/managerial notifications
  - Stakeholder notification or consultation
  - Training and outreach
  - Hiring staff or restructuring units
  - Understanding/agreement concerning roles and responsibilities for the RBS scheme processes
  - Alpha or beta testing of sampling calculators or data system processes
- **When** do required steps need completion before implementation?
- **When** will the RBS scheme/program formally start?

5.6.2. Step 8 – Create or update manuals and other documentation

Many NPPOs document entry conditions for plant products and information on inspection protocols in official manuals. These materials may need updating to describe the new RBS scheme. The level of detail will vary depending on what other materials the NPPO will provide to staff and stakeholders. In addition, work instructions or job aids may need to be developed or may need updating. Such documents may be used to provide needed details or tips for staff as they carry out specific tasks.

One recommended activity is to create a communications plan for the RBS program in general and for the group managing the RBS program. This is different than the plan discussed earlier to communicate inspection outcomes to stakeholders. The scope of the group communications plan is broader. The goal of this communications plan is to identify information needs, which parties need the information, how information delivery will occur, and how often communications will take place.

5.6.3. Step 9 – Complete all other requirements

Examples of tasks and requirements needing completion before implementation of the RBS program were listed above.

It might also be logical to consider a short-term trial of the proposed RBS scheme. Pilot programs are an excellent way to evaluate schemes on a limited basis before the entire program is set in motion. It would allow the NPPO to return to plan adjustments without too much lost time, resources, or effort. A pilot test might also serve as a live introduction of the program to stakeholders. Trial programs can take many forms, from “beta” tests of new sampling tools by inspectors, to simulations, to actual test implementations on parts of pathways (e.g., 2-5 commodities or volunteering importers).

5.6.4. Step 10 – Training and outreach

5.6.4.1. Training

Some NPPO staff might need to be trained to carry out the basic procedures of the new inspection scheme(s). Also, some outreach may be needed for staff or stakeholders who are not directly involved in import processing and inspections but are either indirectly affected by them within the NPPO (or other agencies), or participate within the supply chain in some way, and therefore have economic or other interests in its operation.

Training should provide NPPO staff with the information needed to successfully implement the RBS scheme. We suggest determining the following with as much detail as needed to prepare the NPPO for implementation:

- **Who** needs to learn about the plan?
- **What** does each group need to know about the plan?
- **When** are the relevant parties available for delivery of training/outreach? (How much time is needed?)
- **How** and **where** do we plan to deliver the needed information?

The format and content of training will be specific to the NPPO and the RBS program being implemented. As part of this activity, NPPOs should provide time for familiarization with any required procedures, particularly entering required information and determining sample sizes.

5.6.4.2. Outreach

The goal of outreach is to provide sufficient information about the new program to all interested parties and address their concerns and questions to the extent possible before implementation. External stakeholders, particularly importers and other supply chain entities, likely need information about the program, especially for maximizing the use of incentives. Outreach could include helping NPPO units affected indirectly by the coming implementation to understand the approach and be able to gauge potential impacts, if any. For example, in the United States, NPPO employees working within states may need to know about changes at ports-of-entry that may affect phytosanitary issues and risk in that or nearby states.

NPPOs may use several different approaches for outreach and communicating using multiple means is probably warranted. Outreach approaches may be well established or may need to be developed. Some NPPOs use notification processes that may combine email messages with online postings. For example, the United States uses a process called stakeholder notification (see USDA-APHIS, 2018), while Australia has a very similar process called industry advice notices' (see DAWR, 2016a). Australia has also tested online publishing of longer program explanations for communicating with stakeholders (see DAWR, 2016b). Likewise, EPPO has both a regular publication (EPPO Bulletin) and the EPPO Reporting Service (see EPPO, 2018).

NPPOs sometimes convene special meetings to present important information directly to stakeholders. These allow for two-way communication, but attendance and participation may be limited. Online meetings (webinars) may allow more people to attend. Webinars can be recorded and posted for later viewing. Industry groups or associations might offer ways of getting messages to stakeholders, but these often will not include *every* interested party (or the general public), so should not be the sole means of communication.

Whichever communication means are chosen, it is good practice to publish contact information for program experts or managers, who can answer any remaining questions or address other outstanding issues, or new ones as they arise.

What should be communicated is another important topic. RBS program details will vary by NPPO, but in general the following things may need to be effectively communicated:

- **When** the program begins, indicate how long it will last (if it is temporary), and how often revisions and updates might be considered
- **Who** will be affected by the program
- **What** pathways and commodities are covered, and what materials are specifically eligible and not eligible
- **How** entities or commodities can become eligible for incentives, and whether that requires special procedures or practices
- **What** cost or time savings entities might realize, if eligible for incentives; and
- **How** this program compares to previous processes.

Also, it may be useful to provide offline program documentation which explains the planned program, expected impacts, and provides a justification for the planned design and operation.

5.6.5. Step 11 – Begin the RBS program

Once preparatory tasks have been completed the RBS program can formally begin. Below are some suggestions for making the transition smoother.

First, expect some startup issues. No matter how well the guidelines, procedures, and systems were vetted beforehand, some unforeseen issues or problems are likely to arise. This is especially true if operations are spread out in several or ports-of-entry. Often, this may simply mean clarifying some policies or helping inspectors adjust to the new procedures or tools. To facilitate the startup period, try to ensure that program managers and experts are available to address issues and answer questions, and consider posting advisors at some ports, such as those with very high import volumes.

Second, it may be possible to "jump start" reduced inspection statuses in the program. In the standard acceptance sampling plans discussed above, all commodities typically begin in qualifying inspection (i.e., normal inspection), which might mean significant increases in numbers of inspections or samples taken by inspectors at the start of the program. Over time items would presumably qualify for reduced inspections, reducing the overall inspection effort, but that transition period could take some time and might be too intense or burdensome for staff. To reduce that burden, NPPOs could consider pre-qualifying some items for reduced inspection at the start. Indeed, the NPPO may well have prior inspection data that can be used to determine each pathway's risk status at the start of the program. To accomplish "jump starting", use the most recent data to identify products with the requisite number of sequential accepted lots, according to scheme specifications. These products would begin with reduced inspections status,

which would buffer any inspection increases on other products. If needed, program implementation could even be delayed for a short time to allow such data to be collected. If the new program seems likely not to differ much from the previous program in terms of inspection effort, then perhaps this would not be necessary.

5.7. Maintain inspection scheme

After implementation of the RBS inspection scheme, the remaining tasks are monitoring and reporting on outcomes, and making any needed program adjustments or revisions.

5.7.1. Step 12 – Monitoring the operations and outcomes of the inspection scheme

Monitoring inspection operations and outcomes is an important activity once the RBS scheme has been implemented. Monitoring allows us to see if the program meets, exceeds, or fails to meet the expected outcomes and levels of safeguarding that were previously estimated (Step 4).

The metrics evaluated during monitoring would be mostly unchanged:

- Inspection effort
- Numbers of cleared versus inspected lots
- Total pest detections
- Estimated leakage rates

Other metrics are given in the appendices, and custom calculations for particular programs could also be developed. Importantly, post-implementation data can be compared to historical inspection performance. Agencies can use these analyses to determine if the RBS scheme is performing as expected, or if adjustments may be needed. An important factor to monitor is how many commodity combinations (or entities) demonstrate improved quality over time and, as such, qualify for reduced inspections. If no trend is evident, it may mean that better feedback to importing entities may be needed, or perhaps program adjustments are needed to better incentivize the process.

5.7.2. Step 13 – Adjust inspection schemes as needed

Based on the results of monitoring, NPPOs could consider making minor modifications to the RBS inspection scheme. Examples might include adjusting the reference sampling plan based on a revised choice of *AOQL* (or *C* and *p*, if used), or adjusting the clearance number, *i*. Evaluating the potential impact of any suggested changes (Step 4) is recommended before enacting revisions. The NPPO can determine if the modifications require full review (Step 6), or new training as part of a secondary rollout (Step 7), as well as other activities (Steps 9-10).

5.7.3. Step 14 – (Optional) - Make more complex revisions as needed

Occasionally it might be necessary to make changes to the RBS inspection scheme that go beyond simple and straightforward parameter adjustments. For example, if quality does improve over time as hoped, then adding new levels for reduced inspections might be considered. Other revisions might include:

- Changing verification levels in the MIL-STD-1916 plan
- Adding new *f* levels in a skip-lot sampling plan
- Changing plans entirely, such as from MIL-STD-1916 to SkSP-2
- Converting to a different type of reduced inspections in a custom RBS plan

Because these are more consequential changes, they will require some explanation and justification based on evidence from recent program outcomes. They probably also require new estimates of potential impacts (as in Step 4), perhaps based on information that was not available when the scheme was first developed and implemented. Unlike the minor adjustments considered in the previous step, more complex revisions probably require a complete evaluation and review—Steps 4 and 6—and perhaps some of the implementation and rollout activities presented in Steps 7 to 10, before formally implementing the revised program.

5.8. Implementing ratings-based programs

5.8.1. Background – why choose a ratings-based plan?

An alternative to the acceptance sampling plans discussed above, is to establish a program based on ratings for eligible commodity combinations. In this approach, the NPPO uses collected data on inspection outcomes to rate commodity combinations (e.g., high-risk/low-compliance or low-risk/high-compliance). The ratings determine whether the commodity combination is subjected to reduced, normal, or tightened inspections. This type of program is also called *profiling* in many international jurisdictions (see Clarke et al., 2017). Imports and inspections proceed accordingly with regular updates of the ratings.

In the United States, the USDA-APHIS-PPQ established a ratings-based program called the National Agriculture Release Program (NARP) in 2004. NARP consisted of a list of country-commodity combinations that qualified for a reduced frequency of inspections based on import volume thresholds, the fraction of these imports that were non-conforming, and a review of the identities of the intercepted pests. Ratings were generally updated once per year, but industry stakeholders could request updates on specific commodities at any time that were subject to NPPO analysis and approval.

A ratings-based system permits proactive risk management in circumstances for which changes in biosecurity risk are predictable. For example, there may be a seasonal change in risk that can be detected statistically and justified using knowledge of insect biology. Equally, it may be possible to better predict the future biosecurity risk of entities based on information about their current performance and that of similar entities, by reviewing inspection data (Robinson et al 2015).

Some of the limitations of ratings-based plans include:

- Requirement for enough collected data to allow analysis (and associated delays in implementation)
- Requirement for the establishment of a ratings model
- Requirement to justify and sufficiently explain the workings of the (new, likely) model and rating system to stakeholders
- Requires about twice as many steps (see below) as a cumulative results plan (**Figure 18**) during its creation phase
- Requirement to periodically update ratings based on recent inspection outcomes
- Available data and resulting ratings depend on the time interval between analyses
- Combinations that do not qualify for reduced inspection at the time of implementation *cannot qualify until the next update.*

The major impact of the limitations listed above is directly related to the final point, as it highlights a potentially reduced ability to motivate industry partners to reduce nonconformities. The delays in adjusting the ratings-based approach between importers' instituting improvements and receiving incentives are a limitation. Also, the ratings-based plan may become more about the model or ratings definitions and less about the inspection outcomes, especially if eligibility is dependent on the actions of other entities or if production improvements do not transparently result in greater eligibility. This could put the NPPO on the defensive and allow stakeholders to avoid taking responsibility for improvements or cause them to decide it's not worthwhile. By contrast, the "Send *X* clean shipments" approach to cumulative results plans places the need for improvement squarely on stakeholders.

Currently, dynamic ratings-based programs are rarely used. One exception is a program implemented in 2018 by PPQ at all plant inspection stations for imports of propagative materials (PPQ, 2018a). Effects of the program on inspection effort and longer-term conformity rates are unknown, however. DAWR in Australia uses profiling of the passenger and mail pathways, but uses skip-lot sampling for plant product pathways (Brent, 2016; Robinson et al., 2012a). Interestingly, ratings-based plans are not discussed in modern texts on acceptance sampling (e.g., Schilling and Neubauer, 2017; Shmueli, 2016; Stephens, 2001), nor were they considered during

the formative years of acceptance sampling (see Dodge, 1969). For these reasons, we advise against implementing a ratings-based RBS program, unless a cumulative results plan cannot be used.

Nevertheless, below we briefly discuss the design and implementation of ratings-based RBS programs. The creation and maintenance steps for ratings-based sampling plans are quite different than those discussed earlier. The program implementation step, however, closely follows the description above (see 5.6.1-5), so it will not be repeated here.

5.8.2. Creating ratings-based programs

The overall process is similar to the one described earlier, but the design and maintenance phases have additional steps (**Figure 20**). Implementation is similar to that described above.

5.8.2.1. Step 1 – Identify pathway(s) of interest

This is no different from the text included above (see 5.5.1). However, note that particular attention should be given when defining and tracking eligible commodity combinations, because ratings will eventually need to be calculated for and applied to these commodity combinations.

5.8.2.2. Step 2 – Collect inspection data for analysis

To create ratings for the defined commodity combinations, the NPPO must first collect inspection data for a predetermined time period. Ideally, ongoing data collection processes (see 5.3.1.4) would provide sufficient data, however, some NPPOs may need to begin data collection for the first time. We list this as Step 2 to emphasize the need to start this step as soon as possible, since data is required to progress to other steps.

Because the data will be used to calculate commodity combination ratings, how these inspection outcomes are quantified is more important here than it was before. Often, NPPOs will collect data that allows a reasonably simple estimation of the fraction of non-conforming units (*p;* aka 'action rate' in the USA, or 'quarantine failure rate' in Australia). They may collect less data or no data at all on the fraction of non-conforming units (*d).* Note that estimates of *p* are approximate measures at best because inspection is imperfect and outcomes depend upon both *p* (the true rate, not what is detected) and *d*, which affects the likelihood of detecting non-conforming units. The best measure is likely infestation rate, or number of pests or propagules per inspected unit, but this information will rarely be available because collecting the required data is challenging and costlier than collecting other data (Caton, 2018). Nevertheless, infestation rate allows the prediction of numbers of pests or propagules entering and is the most direct measure of possible pest risk.

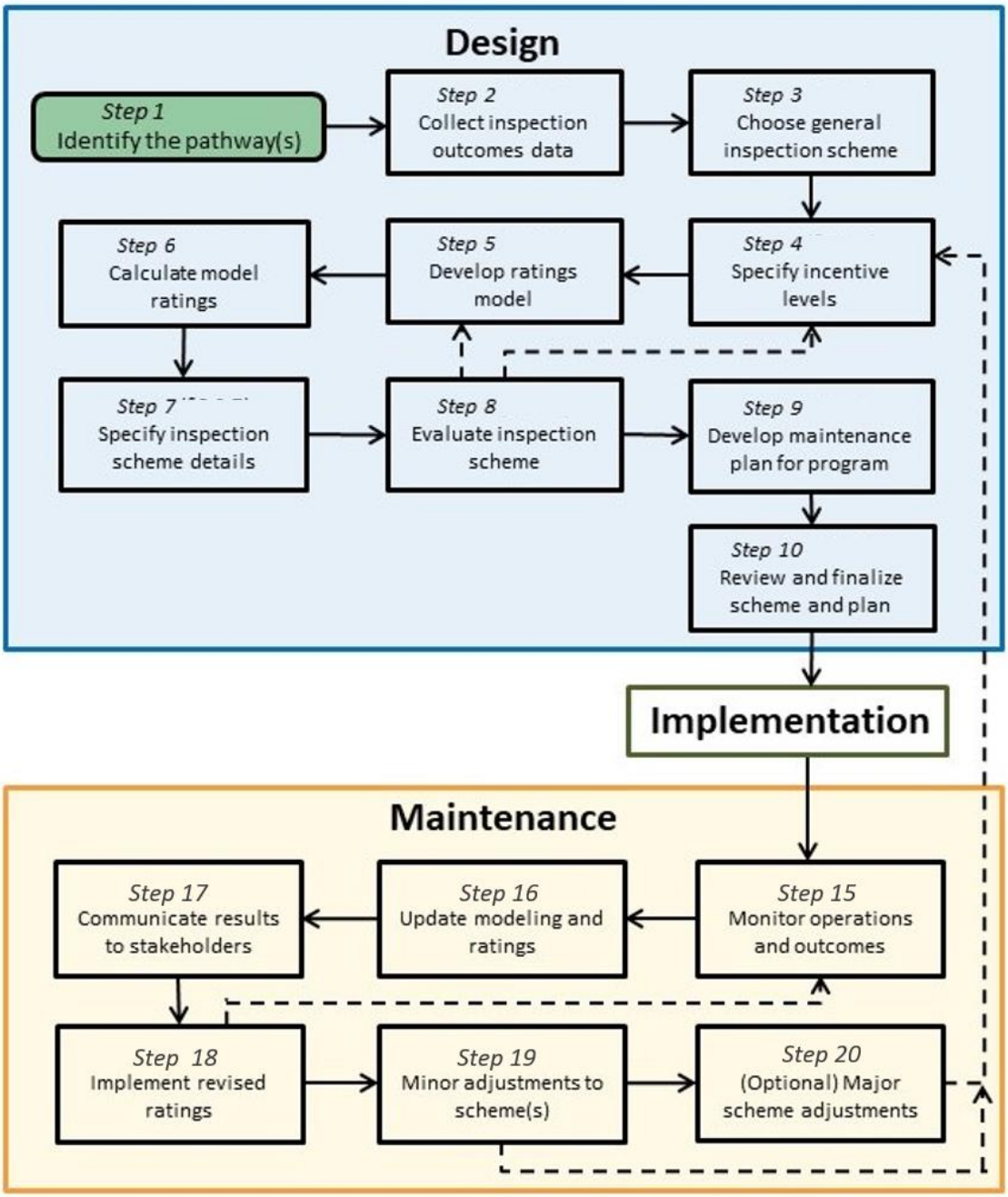


***Figure 20**. Flowchart for the process of designing, implementing, and maintaining a ratings-based RBS inspection scheme. Parentheses list the relevant steps in the text. Dashed lines represent optional paths.*

How much data will be needed depends on the import volume, and also on the method of analysis used. Greater volumes mean shorter data collection times, in general, but if eligible items are specified with more detail (see 5.5.1.2), then data collection times may lengthen to ensure that most combinations have sufficient data. Trade-offs exist between the amount of data collected and the uncertainty associated with ratings.

Each NPPO will need to decide how much data is enough, and below we provide only some general perspectives. Let us assume that the available data has been summarized by lots, and that we will estimate mean *p* and an upper confidence limit using a standard methodology (see Ott and Longnecker, 2001). With these assumptions, the upper confidence limit for *p* for a commodity combination with zero non-conformities (denominator = 0) drops below 0.05 at *N* = 72, below 0.01 at *N* = 368, and below 0.005 at *N* = 597 (a placeholder for the more commonly used 600!).

5.8.2.3. Step 3 – Choose a general inspection scheme

This is similar to the descriptions provided above (see 5.5.2).

5.8.2.4. Step 4 – Specify incentive levels

Completing this step (see 5.5.3.1) before commencing the ratings analysis is important so that analysts know, at least approximately, how many ratings categories are needed. However, the ratings estimation process and results could impact how the inspection scheme is chosen. For example, a very clean pathway might only need two ratings categories - normal and reduced inspections. However, if the pathway contains several combinations with a high mean *p*, the NPPO might want to consider including a category for tightened inspections. Consequently, some iterations between this step and steps 5 and 6 below may occur (which will add to the program development time).

At this point, NPPOs should also determine how to treat commodity combinations that lack sufficient data to make an accurate ratings determination. Possibilities include mandating normal inspection levels or requiring increased levels of inspection so that data are more quickly collected.

5.8.2.5. Step 5 – Develop a ratings model

Calculating the ratings for commodity combinations is the most critical step in the creation of a ratings-based RBS plan. But because ratings-based programs have been seldom used, no consensus exists for a standard modeling approach. In the NARP program discussed above, the standards were 1) exceeding a threshold volume, 2) having no significant pest threats detected on the commodity at any time, and 3) having a mean action rate over the defined time-period below the defined threshold (of 1 percent). Note that uncertainty in the mean action rate was not considered. More recently, the United States NPPO has based some RBS programs on both predicted mean *p* values and an estimate of the associated uncertainty. Robinson *et al.* (2011) argued that risk due to known contamination and risk due to uncertainty could be combined for the purposes of designing a sampling scheme.

Program eligibility. With these above points in mind, NPPOs might decide to separately assess program eligibility. In other words, NPPOs might begin by identifying those commodity combinations for which ratings need to be calculated, based on meeting certain volume threshold, (low) pest risk potential, or other standards.

Modeling methods. Options for a ratings-based model include *empirical* and *model fitting* methods, and the choice has substantial implications for program creation and operation. Empirical methods (see Bolstad and Curran, 2016) use a standardized arithmetical approach for which results change over time only due to the new data outcomes. Model fitting methods (see Clarke et al., 2013, Kim et al., 2018), by contrast, apply a dynamic statistical approach to (ideally) the most recent data, which means that significant model factors and uncertainty thresholds may vary for different time periods. The choice of modelling method has important implications for the resulting RBS scheme (**Table 20**). Whichever approach is chosen, best practices involve estimating the uncertainty associated with the rating.

Table 20. *Factors related to the choice of modeling methodology that impact ratings-based RBS schemes. (Based on Caton, 2018).*

Model Factor	Empirical Modeling	Model Fitting
Specificity	Each country commodity combination estimated separately	All country commodity combinations estimated together at once
Dependency	Results are independent	Results depend on all included country commodity combinations and data
Ratings derivation	Directly determined	Indirectly determined because of dependency
Explicability	Standardized model	Ambiguous and dynamic model and thresholds
Revisions/updates	Single country commodity combinations	All country commodity combinations determined together
Accumulation of data	Possible	Restricted; period-specific data
Update frequency	More frequent (uses historical data)	Less frequent (requires sufficient new data)
Rating factors	Standardized	Dynamic
Uncertainty estimate	Integrated into method [a]	Sometimes requires a separate approach (e.g., simulation)

[a] E.g, true for empirical Bayes

Recent studies suggest that more complicated inspection optimization approaches do not outperform simpler approaches (DeMiguel et al., 2009; Powell, 2015). Moreover, model fitting can seriously constrain program capability by reducing administration flexibility, using historical data, and by perhaps resulting in ever-changing standards (Caton, 2018). Various other considerations come into play depending on model choice (see Decrouez and Robinson, 2013; Robinson et al., 2015). Model fitting methods and their results will likely be more challenging for stakeholders to understand, and therefore result in greater skepticism from users. Consequently,

simpler estimates are recommended, unless clear evidence exists (i.e., strict validation) that more complicated estimates perform markedly better.

We caution against the formulaic application of statistical fitting routines, which can lead to sub-optimal or possibly misleading outcomes. Statistical approaches such as *p*-values and R^2 tell only a portion of the important modeling story. Better outcomes will result when model fitting is aligned to model application. In the development of a model for ratings, model performance should be assessed in light of the intended use of the ratings as well as its statistical characteristics.

Empirical Bayes method. If ratings will be based upon the mean *p* (or other similar metric), then we favor using the empirical Bayes method, which is useful for summarizing a large number of probabilities. Briefly, the method combines the observed proportions into a statistical *prior* distribution (see Appendices F, G). The values of the prior distribution are then combined with the observed proportions, one by one, to come up with bespoke predictions.

The fraction of non-conforming units can be considered as a binomial probability of being non-conforming. A beta distribution (see Vose, 2000) is a formal, quantitative means of describing the estimated binomial probability and the uncertainty around it. Beta distributions are flexible, often nonlinear, curves with a continuous range between 0 and 1. They are described by two shape parameters, typically called *a* and *b* or α and β. Consequently, the estimation method involves finding the best values, *a'* and *b'*, from the new outcomes inspection data and the prior values, a_0 and b_0 (Appendix G). Then an upper confidence interval (e.g., 99 percent limit, p_{99}) can be estimated from *a'* and *b'*.

Compared to using a simple mean *p* (ratio of non-conforming units to total units inspected), the technique improves the analysis by explicitly incorporating the underlying pattern of the data into each estimate. Robinson *et al.* (2015) provide an example in a biosecurity setting along with the R script for implementation, using a slightly different technique for determining the prior distribution.

Model validation. Best practices for both approaches discussed above is to validate the model results. Often this is done with *out-of-sample testing*, which involves reserving some portion of the dataset from the model estimation activity, so that it remains independent. This portion of the dataset can then be used to assess how well the model performs, by comparing real inspection outcomes across the different ratings categories. Some quantitative techniques for model estimation explicitly integrate this process and name it as cross-validation (see Arlot and Celisse, 2010).

Model validation is especially important when data are mostly categorical and sparse, which is common in biosecurity. Results of within-sample model checking can lead to over-inflated

expectations, and potential program failures. Dividing the dataset sample across time is best, to ensure that potential seasonal effects do not overshadow model assessment. If two years of data are available, then a suggested approach is to use data from one year to fit the model and use the date from the second year to assess the quality of the fitted model. The final model – constructed in line with the best approach discovered under cross-validation – can then be fitted using all the data.

Failure of the model during validation likely means restarting this activity. At worst, it could mean collecting more, or different data, further delaying program implementation.

5.8.2.6. Step 6 – Calculate model estimates and assign ratings

Once the modeling method has been chosen and the data has been partitioned into training (model specification) and test (validation) sets, the analysis can proceed. Model fitting can be fairly complicated approach to selecting and specifying the final model. Full consideration of all aspects of the model fitting approach is beyond the scope of this chapter, but see Clarke *et al.* (2017) for examples of model fitting in a biosecurity setting, and Burnham and Anderson (2003) or Johnson and Omland (2004) for more information.

After producing parameter estimates and associated uncertainties for all combinations, the results are used to classify the combinations into discrete groupings by rating. Defining these groupings (or profiles) is a somewhat challenging process (see Linacre, 2002) that NPPOs will need to individually develop and justify. The simplest approach might be to set thresholds based only on the upper confidence limit, but various combinations using both mean p and p_{99} (above) are also possible. Statistical means of finding the best performing grouping are available (see Robinson et al., 2015). Somewhat arbitrary choices that meet specified quality levels (e.g., not likely to be greater than a 0.05 non-conformity rate) can also be used and may vary less than statistically determined thresholds. At a minimum, analysts should evaluate and perhaps tune their ratings performance using the test dataset.

5.8.2.7. Step 7 – Specify details of the inspection scheme

Once the ratings structure has been selected, it should be possible to completely specify the details of the full inspection scheme (see 5.5.3.2 and 5.5.3.3).

5.8.2.8. Step 8 – Evaluate the inspection scheme

This activity is very similar to one described above (see 5.5.4), where it is suggested to at least estimate the inspection effort required for the program, as well as time necessary to perform inspections, and leakage.

Note that an important analytical component for ratings-based programs is the proportion of combinations in each rating category. This will be used to directly estimate the number of arriving lots that receive normal, reduced, or tightened inspections. Normal inspections would include items rated as normal (however this was characterized), and, depending on how the program is formulated, items that may not have qualified because of low import volumes. Estimates for each category will need to be summed to determine overall program totals and expected performance.

5.8.2.9. Step 9 – Develop the program maintenance and feedback plan

This step involves creating the feedback plan, as described above (see 5.5.5), but also requires NPPOs to specify how and when ratings will be updated. The timeline determination will involve tradeoffs between how long data needs to be collected and how often ratings can be updated to allow new commodity combinations to qualify for reduced inspections (or earn increased inspections). More frequent updates would be facilitated by using an empirical modeling approach because it is simpler and requires less new data.

5.8.2.10. Step 10 – Review and finalize the scheme(s) and feedback plan

All nine tasks described above need to be completed before a ratings-based plan can be implemented. This step is similar to that described above (see 5.5.6), except that if the review identifies any required changes in modeling and ratings specifications, significant delays can occur while repeating Steps 6-8.

The establishment of a ratings-based RBS plan ends at Step 10. In the RBS approach discussed earlier the program was finalized by Step 6. A ratings-based plan requires additional activities, and these extra steps may add more time before program implementation can be effected.

5.8.3. Maintenance

Recall that **implementation** of the ratings-based inspection scheme would require the same implementation steps described earlier and as such, will not be presented again here. We will begin the discussion of maintenance activities with Step 15.

5.8.3.1. Step 15 – Monitoring operations and outcomes of the inspection scheme

This step is unchanged from Step 12 described earlier (see 5.7.1). However, an additional metric of interest for ratings-based inspection schemes would be estimating quality by ratings categories. Non-separation of ratings based on their performance would be an indication that the ratings model is not functioning as intended. Ongoing feedback to stakeholders on inspection outcomes should follow the plan created before implementation (see Step 9, 5.8.2.9).

5.8.3.2. Step 16 – Update model estimates and ratings

After a predetermined amount of time set by NPPO experts, the ratings underlying the RBS program will need updating. This activity for ratings-based plans is significant, as it is how the RBS program stays current, rewards improvements in quality with reduced inspections, and increases inspections when quality is suboptimal. Ideally, this activity follows the plan created before implementation (see Step 9, 5.8.2.9). The timeliness for updates is an important consideration – if it is too early then needless work is performed, if it is too late then the ratings will be out of date. Also, changing ratings with regular frequency – for example quarterly may—intentionally or otherwise — capture seasonal changes. However, because current ratings always reflect *past* performance, this update frequency may create a systematic mismatch with *current* products.

The update process will be simplest and most straightforward if the model is unchanged. If model fitting is used, NPPOs can expect model changes to occur, simply because the underlying data will have changed. Any significant model changes, or changes to the threshold levels used to assign ratings, will likely need to be reviewed (see Step 10, 5.8.2.10). Changes to ratings thresholds (Step 6, 5.8.2.6) should be avoided whenever possible, because they can negatively affect program perception and understanding, and, consequently, impact the motivation of stakeholders to comply.

5.8.3.3. Step 17 – Communicate updated results to stakeholders and solicit feedback

This step uses the same communication processes discussed in Step 10 (see 5.8.2.10) related to notification and responses related to program revisions. Notifications without details are not sufficient; stakeholders may need help understanding any ratings or model differences and may want to discuss their data and resulting classification with the NPPO. They may take issue with some aspects of the data, analysis, or results. The delays in updating ratings-based programs adds extra sensitivity and concern to communication of the results from each ratings period.

5.8.3.4. Step 18 – Initiate updates to the inspection scheme

As described earlier (see 5.6.5) any RBS program can formally be restarted after the review is completed and all tasks and adjustments have been completed. Note that the old program, with the previous ratings, will remain active until the switchover occurs. This needs to be clearly understood by all parties.

5.8.3.5. Step 19 – Adjust inspection schemes as needed

Based on the results of monitoring or revisions, NPPOs may need to make minor modifications to the inspection scheme. See examples above (see 5.7.2). Note that this activity is different than just updating ratings because it affects the sampling inspection plan(s). Also, as before, changes may need additional review or secondary roll-out.

5.8.3.6. Step 20 – (Optional) Make more complicated revisions as needed

This optional step allows for more significant changes in the inspection scheme to be developed and enacted (see 5.7.3).

5.9. Conclusions

We hope this Chapter demonstrates how to effectively design RBS programs that both reduce the resources required for inspections and maintain an acceptable level of safeguarding for NPPOs. Herein we suggest that the simplest path for an NPPO to begin using RBS in an inspection program would be to choose a sampling plan from the standard library of possible options. Because these plans have been well vetted from a statistical perspective and have been applied by many different industries, it will likely be simpler to understand them and to justify their use in plant health. NPPOs might want to slightly adapt these RBS plans to meet their needs by adding or deleting (simplifying) some plan features. Creating a new RBS plan from scratch should probably be a last resort, especially for NPPOs attempting to use RBS for the first time.

A lingering question is how much implementation of RBS programs might lead to improvements in phytosanitary compliance and result in improved biosecurity. Even under ideal circumstances and clear incentives, importers may not always choose to adjust their processes to comply with RBS program requirements, for rational economic reasons (Rossiter and Hester, 2017). Within the NARP program in the United States, we documented importing entities trying to qualify for reduced inspections, but in general we did not notice long-term improvements in conformity rates by importing entities. In Australia, recent trial skip-lot sampling programs failed to document cases of importing entities making quality improvements to take advantage of reduced inspection incentives (Brent, 2016). Researchers at CEBRA in Australia continue their work trying to better understand how RBS programs might be redesigned or managed to encourage quality improvement and compliance by importing entities (e.g., Rossiter and Hester, 2017); some older works may also provide useful insights (see Starbird, 2000).

Nevertheless, RBS programs and their statistical underpinnings provide the necessary technical justification and transparency to inspection which is the most used phytosanitary measure around the world. Technical justification is of paramount importance to effect safe and predictable trade of plant products. As such, NPPOs need not wait to begin implementing RBS programs. If they use a sound design to create a sustainable inspection scheme, configure sampling processes to provide trustworthy data, and create incentives for quality improvement using sampling schemes that are understandable to importers, they will improve their inspection processes and meet their fair WTO and IPPC trade obligations as stated in the WTO-SPS Agreement.

Inspection including fruit cutting to detect internal pests.
Source - https://twitter.com/AgrocalidadEC/status/879820151079804928/photo/1

Recording information after completing inspection of Hass avocados.
Source - https://agroexportaciones.com/2021/03/12/piura-ministerio-de-agricultura-certifica-mas-de-850-toneladas-de-palta-hass-para-exportacion/ (Main source: diario El Regional de Piura)

6. BIBLIOGRAPHY

Anonymous, No date. Abbreviated Hypergeometric Tables for Risk-Based Sampling in Commodity Inspection. Unknown, Unknown. 32 pp.

ANSI/ASQ. 1993. American National Standard: Sampling Procedures and Tables for Inspection by Attributes (ANSI/ASQC Z1.4-1993). American National Standards Institute/American Society for Quality (ANSI/ASQ), Milwaukee, WI. 95 pp.

ANSI/ASQ. 1996. American National Standard: An Attribute Skip-Lot Sampling Program (ANSI/ASQC S1 -1996). American National Standards Institute/American Society for Quality (ANSI/ASQ), Milwaukee, WI. 22 pp.

Antila, J., Karhu, T., Mottonen, M., Harkonen, J. and Belt, P. 2008. 'Reducing test costs in electronics mass-production', International Journal of Services and Standards, Vol. 4, No. 4, pp. 393-406.

APHIS- USDA, 2019. Animal and Plant Health Inspection Service- United States Department of Agriculture, Plant Protection Today - Smarter Inspections, Stronger Safeguarding. [En línea] Available at: https://www.aphis.usda.gov/aphis/ourfocus/planthealth/ppq-program-overview/plant-protection-today/articles/rbs

APHIS. 2016. CBP and Trade Automated Interface Requirements: APHIS (CORE) ACE PGA Message Set Implementation Guide. Animal and Plant Health Inspection Service (APHIS), United States Department of Agriculture (USDA), Washington, D.C. 410 pp.

Arlot, S., and A. Celisse. 2010. A survey of cross-validation procedures for model selection. Statistics Surveys 4:40-79.

Arthur, A., S. Zhao, A. P. Robinson, B. Woolcott, E. Perotti, and C. Aston. 2013. Statistical Modelling and Risk-Return Improvements for the Plant Quarantine Pathway. ACERA Project 1206F. Australian Centre of Excellence for Risk Analysis (ACERA), Melbourne, Australia. 127 pp.

Arya, R., Antonisamy, B., & Kumar, S. 2012. Sample size estimation in prevalence studies. The Indian Journal of Pediatrics, 79(11), 1482-1488.

Beale, R., J. Fairbrother, A. Inglis, and D. Trebeck. 2008. One Biosecurity – A Working Partnership. Commonwealth of Australia, Canberra. 244 pp.

Bebbington, M., C. D. Lai, and Govindaraju K, 2003. Continuous Sampling Plans for Markov-Dependent Production Processes under Limited Inspection Capacity. Mathematical and Computer Modelling 38(11):1137-1145

Blackburn, T. M., P. Pysek, S. Bacher, J. T. Carlton, R. P. Duncan, V. Jarosık, J. R. U. Wilson, and D. M. Richardson. 2011. A proposed unified framework for biological invasions. Trends in Ecology and Evolution 26(7):333-339.

Bolstad, W. M., and J. M. Curran. 2016. Bayesian Inference for Binomial Proportion. Pages 149-168 Introduction to Bayesian Statistics, Third Edition. John Wiley & Sons, Inc., USA.

Borghers, E., and P. Wessa. 2017. Statistical Distributions - Beta Distribution - Overview and Example. Office for Research Development and Education. Last accessed February 27, 2017, https://www.xycoon.com/beta.htm.

Brent, C. 2016. Discussion of Risk-Based Sampling Program. Personal communication to B. P. Caton on November 15, 2016, from Biosecurity Assurance, Department of Agriculture and Water Resources, Canberra, Australia.

Burnham, K. P., and D. R. Anderson. 2003. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (Second Edition). Springer-Verlag, New York. 487 pp.

Caton, B. P. 2018. Analysis and data challenges associated with risk-based sampling programs. Pages 54-61 Proceedings International Symposium for Risk-Based Sampling, Baltimore, Maryland, June 26-30, 2017. North American Plant Protection Organization (NAPPO), Inter-American Institute for Cooperation on Agriculture (IICA), United States Department of Agriculture (USDA), Raleigh, NC, USA.

Chen, C., R. S. Epanchin-Niell, and R. G. Haight. 2017. Optimal Inspection of Imports to Prevent Invasive Pest Introduction. Risk Analysis 38(3):603-619.

Clarke, S., A. P. Robinson, M. Chisholm, and G. Hood. 2017. Data Mining Final Report. CEBRA Project 1031A. Centre of Excellence for Biosecurity Risk Analysis (CEBRA), Melbourne, Australia. 156 pp.

Collins, R. D., Jr, K. E. Case, and G. K. Bennett. 1973. The effects of inspection error on single sampling inspection plans. International Journal of Production Research 11(3):289-298.

Daniel, W. W., and C. L. Cross. 2013. Biostatistics: A Foundation for Analysis in the Health Sciences (Tenth Edition). John Wiley & Sons, New York. 960 pp.

DAWR. 2016a. 82-2016 - Importers of peat and selected vegetable seeds may qualify for reduced inspections from 29 August 2016. Australian Government, Department of Agriculture and Water Resources (DAWR). Last accessed December 4, http://www.agriculture.gov.au/import/industry-advice/2016/82-2016.

DAWR. 2016b. Compliance-based inspection trial for peat and selected vegetable seeds. Australian Government, Department of Agriculture and Water Resources (DAWR). Last accessed December 4,

https://web.archive.org/web/20160911095215/http:/www.agriculture.gov.au/import/goods/plant-products/risk-return/trial-peat-vegetable-seeds.

Decrouez, G., and A. Robinson. 2013. Time-Series Models for Border Inspection Data. Risk Analysis 33(12):2142-2153.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? The Review of Financial Studies 22(5):1915–1953.

Department of Defense. 1996. Test Method Standard: DOD Preferred Methods For Acceptance Of Product. MIL-STD-1916. Department of Defense (DOD), Washington, D.C. 28 pp.

Department of Defense. 1999. Department of Defense Handbook: Companion Document to MIL-STD-1916. Department of Defense (DOD), Washington, D.C. 127 pp.

Dodge , H. F. 1943. A Sampling Inspection Plan for Continuous Production. Annals of Mathematical Statistics XIV:264-279.

Dodge , H. F., and H. G. Romig. 1959. Sampling Inspection Tables, Single and Double Sampling (Second Edition). John Wiley & Sons, Ltd., New York. 240 pp.

Dodge, H. F. 1969. Notes on the Evolution of Acceptance Sampling Plans Part I. Journal of Quality Technology 1(2):77-88.

Dodge, H. F. 1977. Keep it simple. Journal of Quality Technology 9(3):102-103.

ECFE. 2005. Recommendation and Guidelines on establishing a Single Window to enhance the efficient exchange of information between trade and government, Recommendation No. 33. Economic Commission For Europe (ECFE), United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT), New York and Geneva, . 33 pp.

Economic Commission For Europe. 2005. Recommendation and Guidelines on establishing a Single Window to enhance the efficient exchange of information between trade and government, Recommendation No. 33. United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT), New York and Geneva. 37 pp.

EPPO. 2006. Phytosanitary procedures: Sampling of consignments for visual phytosanitary inspection. EPPO [European and Mediterranean Plant Protection Organization] Bulletin 36:195-200.

EPPO. 2018. Recommendations to policy makers from Euphresco projects. EPPO Reporting Service no. 11 - 2018; No. 210. European and Mediterranean Plant Protection Organization (EPPO). Last accessed 4 December, 2018, https://gd.eppo.int/reporting/article-6404.

FAO , 2008. ISPM 31 Methodologies for sampling of consignments. [En línea] Available at: https://www.ippc.int/static/media/files/publication/en/2016/01/ISPM_31_2008_En_2015-12-22_PostCPM10_InkAmReformatted.pdf

FAO, 2016. Methodologies for sampling of consignments. International Standard for Phytosanitary Measures No. 31. Rome. Published by FAO on behalf of the Secretariat of the International Plant Protection Convention (IPPC). [Online] Available at: https://www.ippc.int/static/media/files/publication/en/2016/01/ISPM_31_2008_En_2015-12-22_PostCPM10_InkAmReformatted.pdf

Floyd, R., J. Lima, J. deWaard, L. Humble, and R. Hanner. 2010. Common goals: policy implications of DNA barcoding as a protocol for identification of arthropod pests. Biological Invasions 12(9):2947-2954.

Fosgate, G. T. 2009. Practical sample size calculations for surveillance and diagnostic investigations. Journal of Veterinary Diagnostic Investigation 21(1):3-14.

Gould, W. P., 1995. Probability of detecting Caribbean fruit fly (Diptera: Tephritidae) infestations by fruit dissection. Florida Entomologist, 78(3), pp. 502-507.

Griffiths DA. 1973. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of casesof a disease. Biometrics 29:637–48

Hood, Y., J. Sadler, J. Poldy, C. Starkey, and A. P. Robinson. 2019. Biosecurity system reforms and the development of a risk-based surveillance and pathway analysis system for ornamental fish imported into Australia. Preventive Veterinary Medicine 167:159-168.

Hughes G, Madden LV. 1993. Using the beta-binomial distribution to describe aggregated patterns of disease incidence. Phytopathology 83:759–63

IPPC. 2008. International Standards for Phytosanitary Measures, Publication No. 31. Methodologies for Sampling of Consignments. International Plant Protection Convention (IPPC), Food and Agriculture Organization of the United Nations, Rome, Italy. 24 pp.

ISO. 2005. International Standard 2859-3 Sampling procedures for inspection by attributes — Part 3: Skip-lot sampling procedures (Second Edition). International Organization for Standardization (ISO), Geneva, Switzerland. 34 pp.

ISO. 2006. International Standard 3534-2 Statistics–Vocabulary and Symbols—Part 2, Applied Statistics. International Organization for Standardization (ISO), Geneva, Switzerland. 74 pp.

ISO. 2013. International Standard 3951-1 Sampling procedures for inspection by variables — Part 1: Specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection for a single quality characteristic and single AQL (Second Edition). International Organization for Standardization (ISO), Geneva, Switzerland. 19 pp.

ISO. 2017. ISO 28590 Sampling procedures for inspection by attributes — Introduction to the ISO 2859 series of standards for sampling for inspection by attributes. International Organization for Standardization (ISO), Geneva, Switzerland. 11 pp.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. Trends in Ecology & Evolution 19(2):101-108.

Jones, O., A. P. Robinson, M. Shield, and J. Sibley. 2017. The Allocation of Inspection Resources. Pages 1-16 in A. P. Robinson, T. Walshe, M. A. Burgman, and M. Nunn, (eds.). Invasive Species: Risk Assessment and Management. Cambridge University Press, Cambridge, United Kingdom.

Kim, B., S. C. Hong, D. Egger, C. S. Katsar, and R. L. Griffin. 2018. Predictive modeling and categorizing likelihoods of quarantine pest introduction of imported propagative commodities from different countries. Risk Analysis:15.

Lane, et al. 2018. Sample size calculations for phytosanitary testing of small lots of seed. CEBRA report.

Lane, S. E., Cannon, R. M., Arthur, A. D., & Robinson, A. P. (2019). Sample size for inspection intended to manage risk within mixed consignments. NeoBiota, 42, 59–69.

Linacre, J. M. 2002. Optimizing Rating Scale Category Effectiveness. Journal of Applied Measurement 3(1):85-106.

Madden LV, Hughes G, Munkvold GP. 1996. Plant disease incidence: inverse sampling, sequential sampling, and confidence intervals when observed mean incidence is zero. Crop Prot. 15:621– 32

Mamber, S. W., T. Mohr, C. Leathers, E. Mbandi, P. Bronstein, and K. Barlow. 2018. Occurrence of Salmonella in ready-to-eat meat and poultry product samples from US Department of Agriculture–regulated producing establishments. I. Results from the ALLRTE and RTE001 random and risk-based sampling projects, from 2005 to 2012. Journal of Food Protection 81(10):1729-1736.

Minton, G. 1972. Verification Error in Single Sampling Inspection Plans for Processing Survey Data. Journal of the American Statistical Association 67(337):46-54.

Montgomery, D.C. 2009. Introduction to Statistical Quality Control 6th ed.

Ott, R. L., and M. Longnecker. 2001. Introduction to Statistical Methods and Data Analysis (Fifth). Duxbury Press, Belmont, CA. 1051 pp.

Perry, R. L. 1973. Skip-lot sampling plans, Journal of Quality Technology 5(3):123–130.

Powell, M. R. 2015. Risk-Based Sampling: I Don't Want to Weight in Vain. Risk Analysis 35(12):2172-2182.

PPQ. 2015. Agricultural Quarantine Inspection Conversations 1: AQI Reset. Plant Protection and Quarantine (PPQ), Animal and Plant Health Inspection Service, U.S. Department of Agriculture, Riverdale, MD. 2 pp.

PPQ. 2016. Agricultural Quarantine Inspection Conversations 10: Inspection for Information. Plant Protection and Quarantine (PPQ), Animal and Plant Health Inspection Service, U.S. Department of Agriculture, Riverdale, MD. 2 pp.

PPQ. 2018a. APHIS Will Adjust Risk-Based Sampling Procedures at Plant Inspection Stations on September 30, 2018. Plant Protection and Quarantine (PPQ), Animal and PLant Health Inspection Service (APHIS), U.S. Department of Agriculture. Plant Protection and Quarantine (PPQ), Animal and PLant Health Inspection Service (APHIS), U.S. Department of Agriculture, Washington, D.C. Last accessed June 2, 2021, https://content.govdelivery.com/accounts/USDAAPHIS/bulletins/20b5d77.

PPQ. 2018b. Manual for Agricultural Clearance (Second edition). United States Department of Agriculture, Marketing and Regulatory Programs, Animal and Plant Health Inspection Service, Plant Protection and Quarantine (PPQ), Washington, DC. Last accessed http://www.aphis.usda.gov/import_export/plants/manuals/ports/downloads/mac.pdf.

Ramírez Guzmán, M. E., 2017. Predictive phytosanitary model for quarantine pests. Proceedings International Symposium for Risk- Based Sampling, pp. 65-73.

Roberts, J., S. Low-Choy, F. Jarrad, and K. Mengersen. 2015. Common Statistical Distributions Used in Statistical Modelling and Analysis for Biosecurity Surveillance. Pages 348-361 in F. Jarrad, S. Low-Choy, and K. Mengersen, (eds.). Biosecurity Surveillance: Quantitative Approaches. CABI, Wallingford, Oxfordshire, U.K.

Robinson, A. 2018. Inspection Protocol of Department of Agriculture and Water Resources, Australia. Personal communication to B. Caton on July 12, 2018, from

Robinson, A. P., & Hamann, J. D. (2008). Correcting for spatial autocorrelation in sequential sampling. Journal of Applied Ecology, 45(4), 1221–1227.

Robinson, A. P., M. Chisholm, R. Mudford, and R. Maillardet. 2015. Ad hoc Solutions to Estimating Pathway Non-compliance Rates Using Imperfect and Incomplete Information. Pages 167-180 in F. Jarrad, S. Low-Choy, and K. Mengersen, (eds.). Biosecurity Surveillance: Quantitative Approaches. CABI, Boston, MA.

Robinson, A., Cannon & Rob, M. A. B. &., 2011. Allocating surveillance resources to reduce ecological invasions: maximizing detections and information about the threat. Ecological Applications, pp. 1410- 1417.

Robinson, A., J. Bell, B. Woolcott, and E. Perotti. 2012a. AQIS Quarantine Operations Risk Return, ACERA 1001 Study J: Imported Plant-Product Pathways, Final Report. Australian Centre of Excellence for Risk Analysis (ACERA), Melbourne, Australia. 100 pp.

Robinson, A., M. Burgman, and R. Cannon. 2011. Allocating surveillance resources to reduce ecological invasions: maximizing detections and information about the threat. Ecological Applications 21(4):1410–1417.

Robinson, A., R. Cannon, and S. Goldie. 2012b. DAFF Biosecurity Quarantine Operations: Risk-Based Approach, ACERA 1001 Study H Overview of Case Studies. Australian Centre of Excellence for Risk Analysis (ACERA) and Department of Agriculture, Fisheries and Forestry (DAFF), Australia, Melbourne, Australia. 29 pp.

Rossiter, A., and S. M. Hester. 2017. Designing Biosecurity Inspection Regimes to Account for Stakeholder Incentives: An Inspection Game Approach. Economic Record 93(301):277-301.

Rossiter, A., S. Hester, C. Aston, J. Sibley, G. Stoneham, and F. Woodhams. 2016. CEBRA Project 1304C: Incentives for Importer Choices, Final Report. Centre of Excellence in Biosecurity Risk Analysis (CEBRA), Melbourne, Australia. 133 pp.

Schilling, E. G., and D. V. Neubauer. 2017. Acceptance Sampling in Quality Control (3rd). Chapman and Hall/CRC, New York. 882 pp.

Schilling, E., 1982. Acceptance sampling in quality control. ASQC Quality Press. United States of America: s.n.

Shmueli, G. 2016. Practical Acceptance Sampling: A Hands-On Guide (Second Edition). Axelrod Schnall Publishers, USA. 106 pp.

Springborn, M. R., A. R. Lindsay, and R. S. Epanchin-Niell. 2018. Risk-based Inspection: Setting policy parameters to harness enforcement leverage. Pages 48-53 Proceedings International Symposium for Risk-Based Sampling, Baltimore, Maryland, June 26-30, 2017. North American Plant Protection organization (NAPPO), Inter-American Institute for Cooperation on Agriculture (IICA), United States Department of Agriculture (USDA), Raleigh, NC, USA.

Squeglia, N. L. 2008. Zero Acceptance Number Sampling Plans (Fifth Edition). ASQ Quality Press, Milwaukee, Wisconsin. 35 pp.

Starbird, S. A. 2000. Designing Food Safety Regulations: The Effect of Inspection Policy and Penalties for Noncompliance on Food Processor Behavior. Journal of Agricultural and Resource Economics 25(2):616-635.

Stephens, K. S. 1995. Volume 4: How to Perform Skip-Lot and Chain Sampling (Second Edition). ASQC Quality Press, Milwaukee, WI. 64 pp.

Stephens, K. S. 2001. The Handbook of Applied Acceptance Sampling: Plans, Principles, and Procedures. ASQ Quality Press, Milwaukee, WI. 538 pp.

United Nations. 2011. Key Factors in Establishing Single Windows for Handling Import/Export Procedures and Formalities: Trade Facilitation and the Single Window. United Nations Economic and Social Commission for Western Asia (Escwa), New York. 89 pp.

USDA-APHIS. 2018. APHIS Will Adjust Risk-Based Sampling Procedures at Plant Inspection Stations on September 30, 2018. United States Department of Agriculture (USDA), Animal and Plant Health Inspection Service (APHIS), Washington, D.C. Last accessed December 12, https://content.govdelivery.com/accounts/USDAAPHIS/bulletins/20b5d77.

Venette, R. C., Moon, R. D. & Hutchison, W. D., 2002. Strategies and Statistics of Sampling for Rare Individuals. Annual Review of Entomology, pp. 143-174.

Vose, D. 2000. Risk Analysis: A Quantitative Guide (2nd). John Wiley & Sons, Ltd., New York. 418 pp.

Wan, H., X. Xu, and T. Ni. 2013. The incentive effect of acceptance sampling plans in a supply chain with endogenous product quality. Naval Research Logistics 60(2):111-124.

Williamson, M., and A. Fitter. 1996. The varying success of invaders. Ecology 77:1661-1666.

Wolter, K. M., 1985. Introduction to Variance Estimation. Springer-Verlag, New York.

WTO. 2014. The Trade Facilitation Agreement: An overview. World Trade Organization. Last accessed https://www.wto.org/english/tratop_e/tradfa_e/tradfatheagreement_e.htm.

Yamamura, K. 1995. Estimation of the Pest Prevention Ability of the Import Plant Quarantine in Japan

Yamamura, K., H. Katsumata, J. Yoshioka, T. Yuda, and K. Kasugai. 2016. Sampling inspection to prevent the invasion of alien pests: statistical theory of import plant quarantine systems in Japan. Population Ecology 58:63-80.

Yamamura, K., Katsumata, H., Yoshioka, J., Yuda, T., & Kasugai, K. (2015). Sampling inspection to prevent the invasion of alien pests: statistical theory of import plant quarantine systems in Japan. Population Ecology, 58(1), 63–80.

7. APPENDICES

Appendix A — Sample size calculations

Binomial approximation of hypergeometric distribution

The function for sample size, *n*, based on the binomial distribution is (Fosgate, 2009):

$$n = \log(1 - C) / \log(1 - p) \quad \text{[Eqn. A1]}$$

where *C* is the confidence level and *p* is the acceptable nonconformity rate (or prevalence level worth detecting).[16]

Hypergeometric distribution

The hypergeometric function for sample size is (Fosgate, 2009):

$$n = [1 - (1 - C)^{1/D}] \times [N - ((D - 1) / 2)] \quad \text{[Eqn. A2]}$$

where *D* is the expected number of defective units in the population, and *N* is the population size (total quantity of units). In practical terms, estimate mean *D* as follows:

$$D = N \times d \quad \text{[Eqn. A3]}$$

where *d* is the proportion of defective units. If unknown, one might assume $d = p$.

For both hypergeometric equations above, rounding to the nearest integer may sometimes be necessary. For example, $N = 999$ and $p = 0.015$ (assuming $d = p$) gives $D = 14.9$. A fractional unit is likely not possible (e.g., fruit, seeds, plants), so that value should be rounded to 15 for use in Eqn. A2. Likewise, $N = 999$, $D = 15$, and $C = 0.95$ gives $n = 179.6$. Calculated values of *n* should always be rounded *up* to the next integer, to avoid undersampling (Fosgate, 2009).

Sample size as a function of AOQL

Based on an equation for calculating *AOQL* from *n*, published by Dodge and Romig (1959), the following function determines *n* from desired *AOQL* for single sampling plans (Stephens, 2001):

$$n = (y \times N) / [(AOQL \times N) + y] \quad \text{[Eqn. A4]}$$

where *y* is a function of acceptance number, *c*. The value of *y* for $c = 0$ is 0.3679. Some example values are shown for a lot size of 1000 below (Table A1). Note that there is no simple function for relating *AOQ* to *n*, because *AOQ* varies with *p*.

To find *AOQL* (technically, $AOQL_1$) given *n*, the function is:

16 This equation can be rearranged to find any one value if the other two are known:
$C = 1 - 10\text{^}(n \times \log(1 - r))$
$r = 1 - 10\text{^}([\log(1 - C)] / n)$

$$AOQL_1 = y \times [(1/n) - (1/N)] \quad \text{[Eqn. A5]}$$

Table A1. Sample sizes, *n*, as a function of *AOQL* (Eqn. A4) for a lot size, *N*, of 1000.

AOQL	**Sample size, *n***		***AOQL***	**Sample size, *n***	
	Calculated	**Rounded up**		**Calculated**	**Rounded up**
0.001	269.0	269	0.035	10.4	11
0.002	155.4	156	0.040	9.1	10
0.003	109.2	110	0.045	8.1	9
0.004	84.2	85	0.050	7.3	8
0.005	68.5	69	0.055	6.6	7
0.006	57.8	58	0.060	6.1	7
0.007	49.9	50	0.065	5.6	6
0.008	44.0	44	0.070	5.2	6
0.009	39.3	40	0.075	4.9	5
0.010	35.5	36	0.080	4.6	5
0.015	23.9	24	0.085	4.3	5
0.020	18.1	19	0.090	4.1	5
0.025	14.5	15	0.095	3.9	4
0.030	12.1	13	0.100	3.7	4

Appendix B — Probabilities of acceptance (hypergeometric)

Probability of acceptance of a lot from hypergeometric distribution

Use the defined Excel function as follows:

$$P_a = \text{HYPGEOM.DIST}(x,n,D,N,\text{FALSE}) \quad \text{[B1]}$$

where x is the number of defectives found (= 0 if determining baseline probability of acceptance), n = sample size, D = expected (mean) number of defectives in the lot, N = lot size, and 'false' indicates not to return a cumulative value.

Probability of acceptance adjusted by fraction nonconforming

The standard calculation for P_a above, when applied to a series of incoming lots, assumes every lot has D defectives in it, i.e., $p = 1$ (fraction nonconforming). That may not always be true, however. If $p < 1.0$, then P_a needs to be adjusted before being applied to a series of lots. The basic idea is to reduce the fraction of lots rejected, P_r, by the proportion of lots that are conforming (i.e., have zero defectives). Adjust it as follows:

1. Baseline probability of rejection of lot: $P_r = 1 - P_a$ [B2]
2. P_r that is truly nonconforming on average: $P_{r\text{-adj}} = P_r \times p$ [B3]
3. Adjusted P_a: $P_{a\text{-adj}} = 1 - P_{r\text{-adj}}$ [B4]

Appendix C — General evaluation metrics

Mean number of lots inspected during standard (qualifying) inspection:

$$U_Q = (1 - P_a^i) / (P_a^i \times (1 - P_a)) \quad [C1]$$

where P_a is the probability of acceptance under the reference plan, and i is the clearance interval. This value should be rounded up to the nearest integer. It estimates how many lots must be inspected, on average, before the lots become eligible for reduced inspection.

Mean number of lots inspected until a rejection:

$$U_{rej} = 1 / P_r = 1 / (1 - P_a) \quad [C2]$$

where P_r is the probability of rejection under the relevant sampling plan (reduced, likely). This value should be rounded up to the nearest integer. It estimates how many lots will be inspected, on average, before one is rejected because of a nonconformity.

Estimating proportions of lots in qualifying and reduced inspections:

1. Calculate U_Q based on P_a for qualifying inspections [Eqn. C1].
2. Calculate U_{rej} based on P_a for reduced inspections [Eqn. C2].
3. The sum of U_Q and U_{rej} is the expected total run length from the start of qualifying inspections to the end of reduced inspections, on average, U_{tot}.
4. The total number of lots arriving, L, divided by U_{tot} is the average number of switches (S) to reduced inspections and back to qualifying inspections over the entire series of arriving lots. This can be rounded to the nearest integer.
5. The product of U_Q and S is the total number of lots in qualifying inspections, L_Q. L_Q divided by L is the associated proportion.
6. The difference between L and L_Q is the number of lots in reduced inspections, L_{red}. The difference between 1 and the proportion of qualifying lots is the proportion of lots in reduced inspections.

Appendix D — Evaluation metrics for reduced intensity schemes

D1. Total samples taken

The following describes one general approach for estimating the total numbers of samples taken in the proposed inspection scheme and without the RBS program.

1. Estimate the total number of lots (L) that are expected for this commodity or pathway.
2. The product of L and n_{Norm} is the total number of samples taken without RBS.
3. Determine or estimate P_a values for both the qualifying ($P_{a\text{-}Q}$) and reduced ($P_{a\text{-}Red}$) inspection phases, which depend on n, d, N, and, if considered, p (Appendix B).

4. Estimate the proportion of lots which will be inspected during qualifying inspections and reduced inspections. This estimate depends on mean lots inspected during qualifying inspection to reach *i* cleared sequential lots with $P_{a\text{-}Q}$, and upon the mean run length during reduced inspection with $P_{a\text{-}Red}$ (Appendix C).
5. Using the proportions determined above, find the number of lots in qualifying inspection (L_Q) and in reduced inspection (L_{Red}). Multiply each by n (n_{Norm} or n_{Red}) to determine the total number of samples taken for each.
6. The sum of those numbers is the total number of samples taken under RBS.

D2. Total time taken

Some values estimated above are reused here. The approach is as follows:

1. Estimate how long each relevant step in the inspection process should take under both qualifying and reduced inspections (e.g., Table 9), and the total time for each in person-hours (h_Q and h_{Red}).
2. The product of h_Q and L (total lots) is the estimated time for inspections *without* RBS.
3. The product of h_Q and L_Q is the estimated time for qualifying inspections.
4. The product of h_{Red} and L_{Red} is the estimated time for reduced inspections.
5. The sum of those quantities is the total time for inspections under the RBS program.

D3. Leakage

Some values estimated above are reused here. The approach is as follows:

1. From d and N, estimate the mean number of defective units, D, in a nonconforming lot.
2. Multiply L_Q and $P_{a\text{-}Q}$, and then L_{Red} and $P_{a\text{-}Red}$, to determine the number of lots accepted in qualifying and reduced inspections. Sum these to find the total lots accepted (L_A).
3. Multiply L_A by p to estimate total accepted nonconforming lots ($L_{A\text{-}NC}$).
4. Multiply $L_{A\text{-}NC}$ and D to estimate the total number of defective units accepted.
5. For comparison to the non-RBS program, perform the same calculations on all lots but with no reduced inspections (i.e., $P_{a\text{-}Q}$ only).

Appendix E — Evaluation metrics for reduced frequency schemes

E1. Total samples taken

The approach is as follows:

1. Estimate the total number of lots that are expected for this commodity or pathway (L).
2. The product of n_{Norm} and L is the total number of samples taken without RBS.
3. Determine or estimate P_a values for both the qualifying ($P_{a\text{-}Q}$) and reduced ($P_{a\text{-}Red}$) inspection phases, which depend on n, d, N, and, if considered, p (Appendix B).

4. Estimate the proportions of lots inspected during qualifying and reduced inspections. These depend on mean lots inspected during qualifying inspection to reach *i*, and upon the mean run length during reduced inspection (a function of $P_{a\text{-}Red}$) (Appendix C).
5. The number of lots inspected during qualifying (L_Q) is the product of *L* and the proportion of qualifying inspections. The number of lots subject to reduced inspections, L_{Red}, is the difference between *L* and L_Q.
6. The product of L_Q and n_{Norm} is the number of samples taken in qualifying inspections (m_Q).
7. The estimated number of lots *inspected* under reduced inspection ($L_{I\text{-}Red}$) is the product of L_{Red} and *f*, the mean proportion of those lots which get inspected. The number of lots cleared without inspection (L_C) is *L* minus both L_Q and $L_{I\text{-}Red}$. [Total lots inspected, $L_{I\text{-}tot}$, is *L* minus L_C.]
8. The product of $L_{I\text{-}Red}$ and n_{Norm} is the number of samples taken in reduced inspections (m_{Red}).
9. The sum of m_Q and m_{Red} is the total number of samples taken under RBS (m_{Tot}).

E2. Total time taken

The approach is as follows:

1. Estimate how long each relevant step in the inspection process should take under inspections, and the total time (h_I), and how long each step takes when lots are only being cleared without inspection, and the total time (h_C) (Table 9).
2. The product of h_I and *L* (total lots) is the estimated time for inspections without RBS.
3. The product of h_I and $L_{I\text{-}tot}$ is the estimated time for RBS inspections.
4. The product of h_C and L_C is the estimated time taken in the RBS program to clear lots without inspection.
5. The sum of those quantities is the total time for inspections and clearances under the RBS program.

E3. Leakage

The approach is as follows:

1. The product of *d* and *N* is the mean number of defective units, *D*, in a nonconforming lot.
2. The product of *L* and P_a (and *p*, if $p < 1$) is the estimated number of nonconforming lots accepted without RBS ($L_{NC\text{-}}$).
3. The product of $L_{NC\text{-}}$ and *D* estimates the total defective units entering without RBS.
4. The product of P_a and $L_{I\text{-}tot}$ (and *p*, if $p < 1$) is the total number of nonconforming lots accepted amongst lots that were inspected ($L_{NC\text{-}I}$) under RBS.
5. The number of nonconforming lots amongst the cleared lots ($L_{NC\text{-}C}$) is equal to L_C if $p = 1$, or is the product of L_C and *p*, if $p < 1$.
6. The sum of $L_{NC\text{-}I}$ and $L_{NC\text{-}C}$ is the total number of nonconforming lots accepted or cleared ($L_{NC\text{-}tot}$).
7. The product of $L_{NC\text{-}tot}$ and *D* is the estimated total number of defective units entering.

Appendix F — Equations for Empirical Bayes Method Applied to Beta Distributions

The expression for the distribution of the probability of infestation based on sample data and using the beta distribution is (Bolstad and Curran, 2016):

$$p_{inf} = \text{Beta}(a', b') \quad [F1]$$

where a' and b' are the posterior or updated parameter estimates. The expressions for a' and b' are:

$$a' = a_0 + N_{NC} \quad [F2]$$
$$b' = b_0 + N_{insp} - N_{NC} \quad [F3]$$

where a_0 and b_0 are prior parameter values, N_{inf} is the number of nonconforming (infested) lots, and N_{insp} is the number of inspected lots.

The equation for the mean (μ_{beta}) of the beta distribution is as follows (Vose, 2000):

$$\mu_{beta} = a' / (a' + b') \quad [F6]$$

The equation for a percentile (p_x) for the beta is as follows:

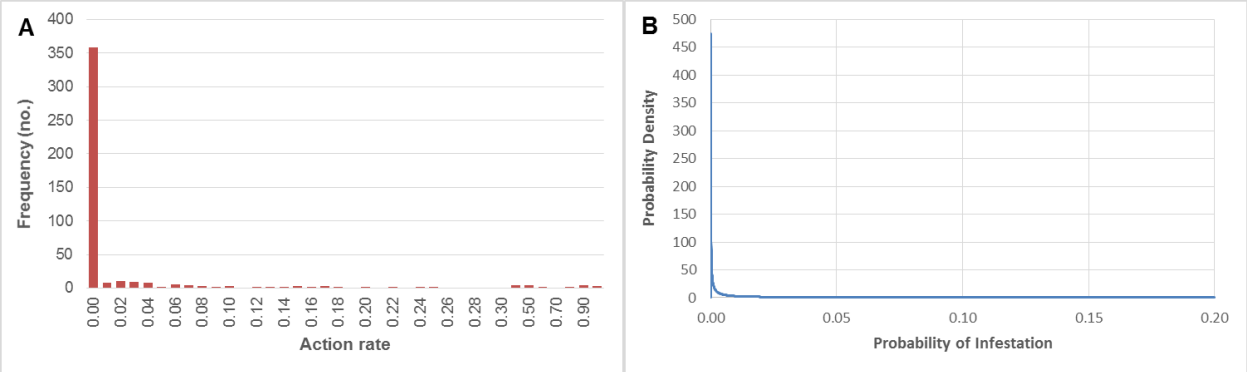
$$p_x = \text{Beta.Inv}(x, a'_i, b'_i) \quad [F7]$$

where Beta.Inv is a function for the inverse of the beta distribution in Excel, and x is the percentile being evaluated (i.e., x = 0.99 for 99th percentile).

Appendix G — Estimating a Bayesian Prior for a Beta Distribution from Overall Likelihood

Here we give an example of estimating a Bayesian prior using the inspection outcomes data for all commodity combinations for which specific distributions will be calculated. The data are calculated values of p (fraction nonconforming or action rate) for 451 commodity combinations. A histogram of values for these combinations indicates that 0 is the most likely action rate (**Figure G1**). For that histogram, the mean ($\hat{p}$) was 0.0391, and the variance (σ) was 0.02025.

Figure G1. (A) Histogram of nonconformity rates for 451 commodity combinations. (B) Prior beta distribution for the probability of being nonconforming based on the mean and variance of (A).



Estimate a_0 and b_0 from the histogram data as follows (after Borghers and Wessa, 2017):

$$a_0 = \hat{p}^2 \times [((1 - \hat{p}) / \sigma) - (1 / \hat{p})] \quad [G1]$$

$$b_0 = a_0 \times [(1 / \hat{p}) - 1] \quad [G2]$$

This gave a_0 = 0.033 and b_0 = 0.822 (Fig. 1b). This prior (Fig. G1B) indicates that a randomly chosen combination is likely to be much closer to 0 than 1.

1. Calculate $P_{a\text{-}Q}$ for qualifying (standard) inspections, which depends on mean lot size (N_{mn}), n_Q (normal sample size), and the expected (mean) number of defective units per lot (D). D is estimated as the product of N_{mn} and d, the proportion of units that are defective.
2. Calculate U_Q, the mean number of lots inspected during qualifying inspection [Eqn. C1]. If $U < i$, then reduced inspection will not often be achieved, and the inspection scheme parameters should probably be adjusted (go to Step 3).
3. Calculate $P_{a\text{-}Red}$ for reduced inspections, as above but using reduced sample size, n_{Red}.
4. Calculate U_{rej}, the mean number of lots inspected until one is rejected [Eqn. C2].
5. Divide the total number of incoming lots, L_{tot}, by the sum of U_Q and U_{rej}, which is the mean number of lots from the start of qualifying to the first rejection. This estimates how many nonconformities will be found over the course of L_{tot}, or the number of switches back to qualifying inspection from reduced inspection (S).
6. Find the number of lots inspected under qualifying inspection, L_Q, as $i + i \times S$, and the total number of samples taken in those lots, m_Q, as $L_Q \times n_Q$.
7. Find the number of lots inspected under reduced inspection, L_{Red}, as $L_{tot} - L_{Red}$, and the total number of samples taken in those lots, m_{Red}, as $L_{Red} \times n_{Red}$.
8. The total number of lots inspected in the proposed scheme, L_I, is $L_Q + L_{Red}$. Compare that to the total number of lots inspected without RBS, which is the product of L_{tot} and n.

Appendix H - Sampling fraction (f) for some values of p, k = 4 and rejection reliability of 95, 80 and 50%.

95%				80%			50%		
p	j	i	f	J	i	F	j	i	f
0.03	50	300	0.0200000000	239	300	0.0041824530	1,000	300	0.0010000000
0.03	219	350	0.0045613741	1,098	350	0.0009111113			
0.03	1,000	400	0.0010000000	5,031	400	0.0001987598			
0.03	5,317	450	0.0001880710	22,867	450	0.0000437303			
0.03	22,215	500	0.0000450155						
0.03	100,000	550	0.0000100000						
0.05	1	100	0.7000000000	7	100	0.1522038695	26	100	0.0380641080
0.05	18	150	0.0553761770	86	150	0.0116628776	344	150	0.0029086301
0.05	238	200	0.0042021951	1,119	200	0.0008938201	4,453	200	0.0002245923
0.05	2,849	250	0.0003509463	14,510	250	0.0000689176	58,040	250	0.0000172295
0.08	1	50	0.9999000000	4	50	0.2845133226	14	50	0.0710919675
0.08	2	60	0.5786021104	8	60	0.1220658007	33	60	0.0306704032
0.08	4	70	0.2524595379	19	70	0.0531799248	76	70	0.0130843571
0.08	9	80	0.1100177222	44	80	0.0229784444	176	80	0.0056920861
0.08	21	90	0.0473820904	100	90	0.0100000000	404	90	0.0024747045

Appendix I- Montecarlo estimate of i pest-free shipments for various values of p1.

p_1	Estimation of i
0.001	9,026
0.002	4,031
0.003	2,515
0.004	1,800
0.005	1,389
0.006	1,123
0.007	939
0.008	804
0.009	701
0.01	620
0.011	555
0.012	502
0.013	457
0.014	419
0.015	***387***
0.016	359
0.017	335
0.018	313
0.019	294
0.02	277
0.021	262
0.022	248
0.023	235
0.024	224
0.025	214
0.026	204
0.027	195
0.028	187
0.029	180
0.03	173
0.031	166
0.032	160
0.033	155
0.034	149
0.035	144
0.036	140

p_1	Estimation of i
0.037	135
0.038	131
0.039	127
0.04	124
0.041	120
0.042	117
0.043	114
0.044	111
0.045	108
0.046	105
0.047	103
0.048	100
0.049	98
0.05	95
0.051	93
0.052	91
0.053	89
0.054	87
0.055	85
0.056	84
0.057	82
0.058	80
0.059	79
0.06	77
0.061	76
0.062	74
0.063	73
0.064	72
0.065	70
0.066	69
0.067	68
0.068	67
0.069	66
0.07	65
0.071	63
0.072	62

p_1	Estimation of i
0.073	61
0.074	60
0.075	60
0.076	59
0.077	58
0.078	57
0.079	56
0.08	55
0.081	54
0.082	54
0.083	53
0.084	52
0.085	51
0.086	51
0.087	50
0.088	49
0.089	49
0.09	48
0.091	48
0.092	47
0.093	46
0.094	46
0.095	45
0.096	45
0.097	44
0.098	44
0.099	43
0.1	43
0.101	42
0.102	42
0.103	41
0.104	41
0.105	40
0.106	40
0.107	39
0.108	39
0.109	39
0.11	38