United States Department of Agriculture

Development and Application of the Risk-Based Sampling Statistical Model Used in Plant Inspection Stations

United States Department of Agriculture

Plant Protection and Quarantine

June 2026

Agency Contact:
PPQ Risk-Based Sampling Team
Plant Protection and Quarantine
Animal and Plant Health Inspection Service
United States Department of Agriculture

Introduction

In 2016 the Plant Protection and Quarantine (PPQ) Risk-Based Sampling (RBS) Program managers directed plant inspection stations (PIS) to transition away from 2% non-random sampling and implement a randomized and statistically valid sampling method. During this time, data was collected and used to build an RBS statistical model that would support the statistically valid randomized sampling. The focus was to move away from a fixed, non-random proportion of sampling units to using randomly selected sampling units based on a hypergeometric probability distribution, which is valid for modeling the likelihood of success out of a number of "trials" or units inspected during an inspection. The change to randomized sampling and development of the RBS statistical model made major improvements to propagative plant material inspections at PIS including:

- Consistent and technically defensible risk management because detection probability no longer varies with shipment size
- Decision-making based on a published, statistically robust approach
- Operationally feasible inspection process, particularly smaller samples sizes for larger shipments, that is fair to trade partners
- Dynamic risk management by updating compliance levels regularly with current data

If the sample rate is fixed at 2% rather than determined by the tolerable level of infestation, the detection levels from shipment to shipment are inconsistent. If every shipment contained the same number of units, percentage sampling would give us a consistent level of detection; however, lot sizes are variable. For example, two units sampled from a lot of 100 likely results in detection only when the rate of infestation is very high, whereas 2,000 units sampled from a lot of 100,000 can detect very low rates of infestation. Thus, in percentage sampling the level of detection varies with shipment size.

To manage the risk posed by pests, we need to be able to detect pest infestations at a consistent level, meaning percentage sampling is not a technically defensible risk management tool because in small shipments only large infestations will be detected and large shipments will often be oversampled. To comply with phytosanitary trade agreements, an inspection method is needed that consistently detects pests at a given level of infestation across varying lot sizes, which is achieved by RBS. Risk-based sampling enables choosing the appropriate detection level for the shipment and knowing with a high level of confidence the number of units that need to be inspected to detect whether the shipment is infested or not.

The RBS statistical model is the foundation of the RBS Calculator hosted in the Agricultural Risk Management (ARM) system. The PIS RBS statistical model is scripted in R and run quarterly by the RBS Analytics Team to produce a compliance lookup table that is referenced by the RBS calculator. Compliance levels (high, medium, low, and poor compliance) and their associated infestation detection levels are assigned to each origin and propagative material type (OPM) combination based on thresholds of the probability of a quarantine pest being found in a shipment and the uncertainty around the estimates of the probability. The calculator then provides randomly identified sampling units for PIS personnel to inspect. The assignment of compliance level and thus the inspection intensity (number of sampling units inspected) for OPM combinations is updated quarterly.

This document summarizes the development and application of the RBS statistical model for OPM combinations (Figure 1). The document covers:

- Extracting and analyzing PIS inspection data from the ARM Datamart
- Determining compliance categories using statistical modeling to estimate probability and uncertainty of OPM combinations carrying quarantine insect pests
- Implementing data-driven updates to compliance determinations based on inspection results.

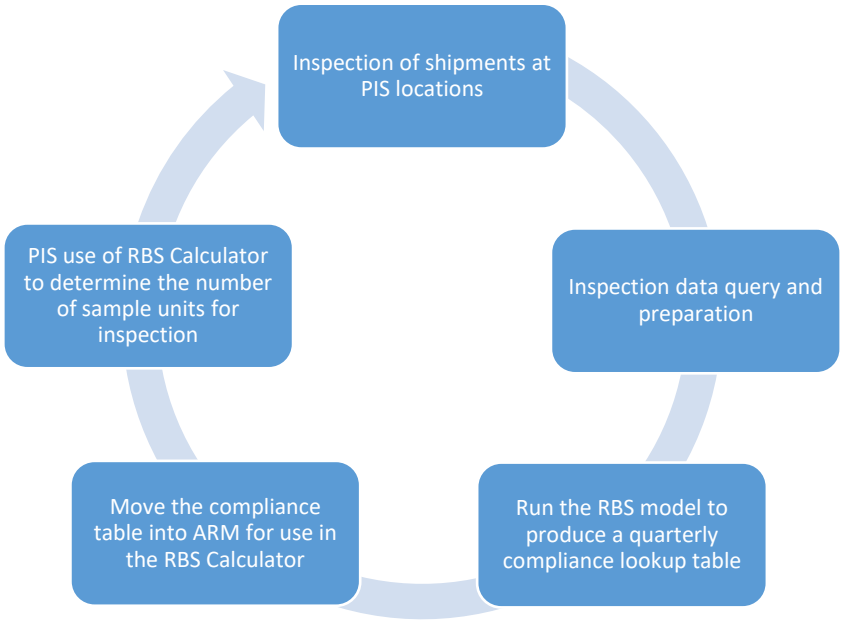


Figure 1. Overview of the risk-based sampling process. Shipments are inspected based on RBS Calculator output and data are collected. Data are queried and prepared, then used in the RBS statistical model to update the compliance levels. Compliance levels are updated in the RBS Calculator in ARM.

Inspection of Shipments at PIS Locations

All plant inspection stations use risk-based sampling. Risk-based sampling is not meant to change the way cargo is held and treated; only the way samples are selected for inspection. After selecting the samples using the RBS Calculator, PIS personnel follow these steps:

1. Remove all contents of the sample unit (e.g., bag, box, or bundle).
2. Look for any contaminants such as Federal noxious weeds, pests, soil, or fomites.
3. Visually inspect the packing material, and if present, the approved growing media, to ensure compliance with USDA regulations.
4. Conduct an overall inspection of the plant material for signs of internal and external feeders and symptoms of disease—being sure to thoroughly inspect the container of the sample unit.
5. Focus inspection on areas of the plant units most likely to be a host to a pest or disease.
6. Use the appropriate inspection methods and tools to conduct focused examinations.
7. Finish inspecting the entire sample unit. If a pest or disease is found during inspection, then the pest is collected and the inspector continues to inspect the full sample.

Inspection Data Query and Preparation

The data used for the RBS statistical model and subsequent quarterly compliance updates are recorded in ARM by PIS personnel as part of the propagative plant material inspection process. The data that are queried from ARM go through an intensive data cleaning and quality control process in preparation for analysis.

The data preparation process includes aggregating unique OPM combinations and calculating an action rate for each combination. Specific disposition codes (Appendix Table 1) are used so that only shipments that were inspected and/or had action taken due to a pest interception are included in the action rate calculation. An action rate is calculated for each OPM combination by dividing the number of actions—represented as actionable disposition codes—by the total number of inspected shipments.

Within an OPM combination, propagative material types include Budwood/Graftwood (scion), Unrooted Cutting, Bulb, Corm, Rhizome, Tuberous Stem Rooted Plant (including grafted), Root cutting, crown, or clump, Tuberous Root, Unrooted Plant, Rooted Cutting (including air layer), Meristem or Callus Tissue Culture (micropropagated/in vitro culture), and Ex-agar. Origin is defined here as the country where the plants were grown, or where the plants from which the plant products are derived were grown. (IPPC, 2016)

Running the RBS Statistical Model to Produce a Quarterly Compliance Lookup Table

At the center of the RBS statistical modelling process is a logistic regression model. The logistic regression model is run each quarter to the past 12 months of data to estimate the probability of quarantine pest interceptions for each OPM combination. The dynamic statistical model is replicated 1,000 times to get a range of plausible probability values which enables quantifying the uncertainty of the probability estimate for each OPM combination. A narrow range of plausible probability values indicates low uncertainty, whereas a wide range of plausible probability values indicates high uncertainty (Figure 2).

To handle uncertainty, a simulation is conducted to obtain the possible range of estimated model coefficients. For each OPM combination, the upper 2.5 and lower 97.5 percentiles are used to simulate action rates. This range represents a possible action rate interval with 95% confidence, providing a measure of uncertainty around the predicted rates. The wider the range, the greater the uncertainty. This range is then used to categorize OPMs into different compliance levels.

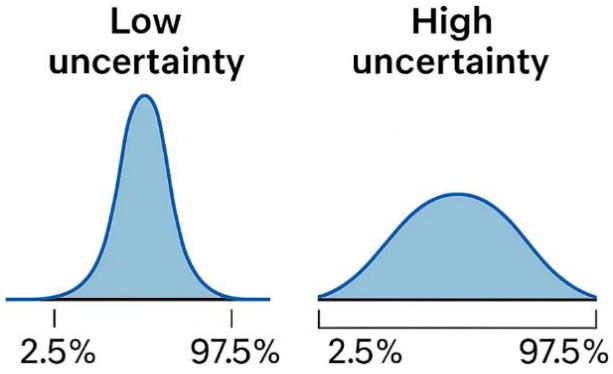


Figure 2. A narrow range of probability values results in low uncertainty whereas a relatively high range of probability values results in high uncertainty.

To find OPM combinations with consistently reliable probabilities of carrying quarantine pests, the lower and upper bounds from simulated probabilities along with their absolute differences are used. The simulated probabilities provide 95% confidence intervals for each OPM combination's probability of carrying pests and the difference between the lower and upper bounds reflects the uncertainty in those estimates. A smaller difference means the probability falls within a narrow range, indicating low uncertainty. The OPMs are then grouped into categories of low or high uncertainty based on the range

of the predicted probabilities. OPM combinations with relatively small differences between the low and high probability values are categorized as 'Low' uncertainty and OPM combinations with relatively large differences between the low and high probability values are categorized as 'High' uncertainty (Figure 3) (Kim et al., 2019).

OPM combinations in the 'low' group have relatively low uncertainty in their action rates compared to those in the 'high' variation group. Threshold values used to group OPMs as low or high uncertainty are determined using simulations and guidance from subject matter experts. The combination of uncertainty and predicted probability of carrying a quarantine pest determines the compliance level of the OPM (Table 1). High compliance is limited OPM combinations in the low variation group, so reduced inspection is applied only to OPM combinations with low uncertainty. The tissue culture PM type is treated as a special case and assigned its own compliance category because of its low action rate. As such, it is subject to very low intensity inspection.

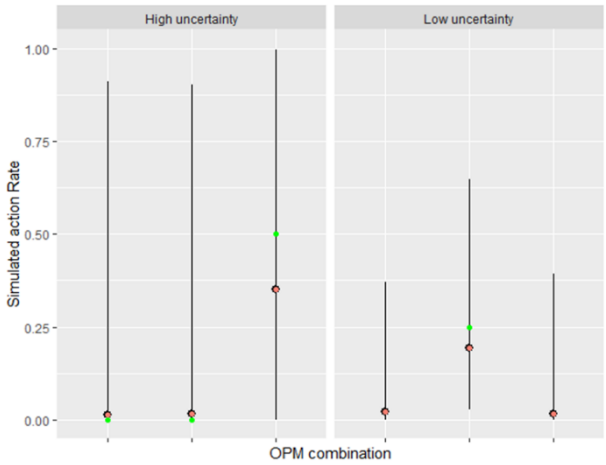


Figure 3. Quarantine action disposition rates of observation (green dots) and simulation (salmon dots) with credible intervals (vertical bars) of country–commodity combinations.

Within a compliance level, all OPM combinations are subject to the same level of inspection. For example, if there were 1000 rooted cuttings each from Guatemala and Honduras and both OPM are in the medium compliance level, then the same number of sampling units of each OPM would be selected for inspection. See the *Application of Compliance in the RBS Calculator* section in the Appendix for more information.

Table 1. Combinations of uncertainty grouping and thresholds of quarantine pest probabilities for assigning compliance levels to OPM combinations.

Uncertainty group	Predicted probability (*P*) threshold of carrying quarantine pests	Compliance level
Low	$P \leq 0.3\%$	High
	$0.3\% < P \leq 3\%$	Medium
	$3\% < P \leq 10\%$	Low
	$P > 10\%$	Poor/Unacceptable
High	$P \leq 3\%$	Medium
	$3\% < P \leq 10\%$	Low
	$P > 10\%$	Poor/ Unacceptable

In Table 2, the maximum sample sizes to detect the selected infestation levels with at least 95% confidence are presented. Using the previous example of Guatemalan and Honduran rooted cuttings, the compliance level is medium, so the maximum sample size to inspect is 13 sampling units to detect an infestation rate of 20% or more with at least 95% confidence.

Table 2. Adjusted detection levels and the maximum number of sampling units for each compliance level. Tissue culture is a special case that is considered above high compliance and so is subject to very low intensity inspection.

Compliance level	Infestation detection level	Maximum sampling units to inspect
High	30%	8
Medium	20%	13
Low	15%	18
Poor	10%	28
Tissue culture	90%	2

Moving the Compliance Table into ARM for Use in the RBS Calculator

The primary output of the RBS statistical model is the compliance level lookup table. Once the RBS Analytics team runs the model, the output lookup file is loaded into the ARM production environment for reference by the RBS Calculator. From there, the PIS personnel will be using the most recently available quarterly compliance for each OPM combination inspected.

Conclusion

The implementation of the RBS statistical model and its use through the RBS Calculator has supported more consistent and defensible inspection activities at PIS locations. By standardizing detection levels across variable shipment sizes, the approach strengthens technical credibility while maintaining operational feasibility for inspectors. Overall workload has remained manageable, with some stations noting reduced sampling demands for larger shipments and clearer expectations for inspection intensity. As PIS personnel have become familiar with the tool and its quarterly updates, acceptance has continued to grow, and the model's routine refresh of compliance levels has improved the utility of inspection data for day-to-day decision making. Collectively, these improvements have enhanced the consistency, transparency, and practical value of the inspection process across PIS operations.

Literature Cited

Fosgate, G.T. (2009), Practical Sample Size Calculations for Surveillance and Diagnostic Investigations. *Journal of Veterinary Diagnostic Investigation*. 21(1):3-14. doi:10.1177/104063870902100102

IPPC. (2016), International Standards for Phytosanitary Measures, Publication No. 5. Glossary of Phytosanitary Terms. *International Plant Protection Convention, Food and Agriculture Organization of the United Nations, Rome, Italy.* 38.

Kim, B., Hong, S.C., Egger, D., Katsar, C.S. and Griffin, R.L. (2019), Predictive Modeling and Categorizing Likelihoods of Quarantine Pest Introduction of Imported Propagative Commodities from Different Countries. *Risk Analysis*. 39: 1382-1396. https://doi.org/10.1111/risa.13252

Appendix A. Disposition codes used

This table provides the disposition codes used to calculate action rates for each origin-propagative material (OPM) type combination in the statistical model and in updates to the quarterly compliance lookup table.

Disposition Code	Disposition Description	Action Taken
DEAP	Destroyed-Actionable Pest	Yes
DEPQ	Destroyed Precautionary	Yes
FUAP	Fumigated-Actionable Pest	Yes
FUPQ	Fumigated Precautionary	Yes
IRMR	Inspected & Released-Meets Requirements	No
IRPD	Inspected & Released-Phyto Discrepancy Corrected	No
OTAP	Other Action Taken-Actionable Pest	Yes
OTCC	Other Action Taken-Carrier Contamination	Yes
OTPQ	Other Action Taken Precautionary	Yes
RXAP	Returned-Actionable Pest	Yes
RXNC	Returned, Non Compliant Wood Packing Material	Yes
RXPQ	Returned Precautionary	Yes

Appendix B. Application of Compliance Levels in the RBS Calculator

Compliance levels of country of origin and PM type combinations were used to assign different sampling efforts depending on OPM compliance. The formula based on hypergeometric sampling was used (Fosgate, 2009):

$$n = \left(1 - \alpha^{\frac{1}{D}}\right) * \left(N - \frac{D-1}{2}\right)$$

where α is 1-confidence, D is the number of sample units requiring action (or detection/infestation level), and N is the total number of sample units inspected (box, baggie, plant unit, etc.). The function code in R is:

```
fosgate.sample1 <- function(box, p, ci){
  ninf <- p*box
  sample.size <- round(((1-(1-ci)^(1/ninf))*(box-(ninf-1)/2)))
  sample.size
}
```

where “box” is the total number of sampling units, “p” is the detection/infestation level, and “ci” is confidence level. Values of p were adjusted to reduce sampling efforts depending on compliance level.