EXPLANATORY DOCUMENT
ON
INTERNATIONAL STANDARD FOR PHYTOSANITARY MEASURES
No. 31
(*METHODOLOGIES FOR SAMPLING OF CONSIGNMENTS*)

Carolyn F. Whyte
Analysis and Profiling Group Manager
MAF Biosecurity New Zealand
Auckland, New Zealand

September 2009

Note: Explanatory documents for International Standards for Phytosanitary Measures (ISPMs) are produced as a result of a decision of the Interim Commission on Phytosanitary Measures in 2004 (reported on paragraph 111 of the report of ICPM-6). They are written to provide supporting information to the standard they refer to and cannot be taken as an official legal interpretation of the IPPC or its related documents, and are produced for public information purposes only. Each document is written by an expert, reviewed by at least two peers (usually from the Expert Working Group concerned), then reviewed by the Standards Committee and the IPPC Secretariat. However, the material presented in explanatory documents remains the opinion of the writer and cannot be interpreted as a decision of the ICPM/CPM. It is hoped that most standards will have one or more explanatory documents associated with them.

PURPOSE OF ISPM No. 31 AND ITS RELATIONSHIP WITH OTHER STANDARDS

The standard *Methodologies for Sampling of Consignments*, ISPM No. 31, was written to outline different types of sampling methods that National Plant Protection Organisations (NPPOs) might use to verify compliance of consignments with phytosanitary requirements. The standard discusses various sampling methods and terminology, and also provides a set of tables of sample sizes for use with the statistical sampling methods outlined in the ISPM. These tables are designed to allow NPPOs to determine appropriate sample sizes for general phytosanitary inspection or testing of consignments ("inspection" and "test" are used as in ISPM No. 5, *Glossary of Phytosanitary Terms*).

ISPM No. 31 was intended to complement ISPM No. 23, *Guidelines for Inspection*, which describes procedures for inspection of consignments at import and export for verification purposes. Other ISPMs also mention sampling for inspection, as a general component of a phytosanitary import regulatory system (ISPM No. 20), for mitigating risk (ISPM Nos. 11, 14) or for verifying that tolerances have not been exceeded (ISPM Nos. 16, 21). This standard may also provide general guidance to NPPOs using ISPM No. 6, *Guidelines for Surveillance*, which mentions statistically-based sampling for surveys of geographical areas.

Other ISPMs that also refer to sampling to determine the pest status of an area are ISPM Nos. 8, 10, 22 and 26.

This document is not a standard, but was written to provide additional explanation around the terms and concepts in ISPM No. 31. Examples are in the context of sampling consignments rather than field sampling, as this is the intent of ISPM No. 31. An appendix giving examples of how to calculate sample sizes in Microsoft Excel is included at the end of this explanatory document.

GENERAL FORM OF ISPM No. 31

The intent of the standard *Methodologies for Sampling of Consignments* is to guide NPPOs in selecting appropriate methods to sample consignments for inspection or testing. It assumes that readers have a basic understanding of sampling and the outcomes they want to achieve. The standard identifies some common objectives for sampling consignments, and outlines the basic requirements for a sampling design, including identification of lots and sample units, choice of a method and determination of sample size. It also describes the statistical sampling parameters, and advantages and disadvantages of several sampling methods. The appendices, although not a formal part of the standard, describe methods for determining appropriate sample sizes, and provide tables of sample sizes for varying parameter values. The headings of this explanatory document match those of the standard, but have been re-arranged in some cases to discuss concepts as they would be used.

CONTENTS OF THE STANDARD

Sampling Objectives

NPPOs may sample consignments for a variety of reasons. The main reason is to determine if a consignment adheres to the criteria for import or export in terms of the presence or absence of pests and disease. Importing NPPOs want to ensure that the commodities they import comply with their import requirements. Exporting NPPOs want to verify that all required conditions are met in order to provide phytosanitary certification and reduce the chance that the products they export will be rejected by the importing country.

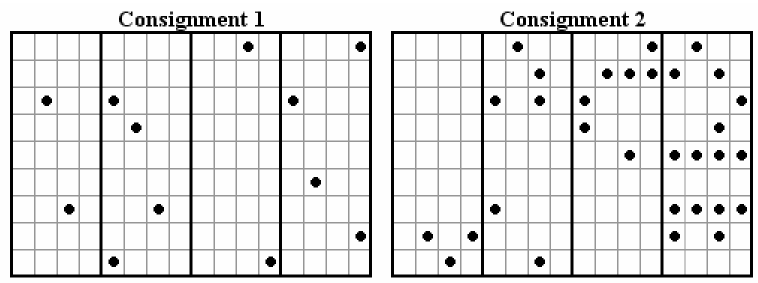
Other reasons for sampling consignments include determining the incidence of infestation in a consignment, determining whether a treatment has been successful, verifying that the consignment matches the description on an import manifest, determining the effectiveness of inspection or treatment procedures and validating compliance with packaging requirements.

It is essential to understand the objective of sampling in order to design an appropriate sampling system. For instance, sampling to provide a specified level of assurance about the pest status of a consignment may use a different method than sampling to determine the proportion of a consignment infested, or to maximise the likelihood of detecting specific pests in a consignment. Some sampling methods may support multiple objectives, but others will not.

Because sampling does not inspect every item, it always involves a degree of error. A statistical sampling method can provide a specified level of assurance that infestation in a consignment does not exceed a particular level (greater than zero). However, a pest-free sample should not be interpreted as providing 100% confidence that the infestation level of a consignment is zero.

Lot Identification

A "lot" is defined in ISPM No. 5 as "a number of units of a single commodity, identifiable by its homogeneity of composition, origin etc., forming part of a consignment". In ISPM No. 31 *Methodologies for Sampling of Consignments*, the lot is the population from which a sample will be drawn. A lot may be an entire consignment or a sub-set of a consignment. In either case, a lot should be reasonably homogenous with regards to pest distribution if a statistical inference is to be made from the sample results.



To illustrate what is meant by homogeneity, assume that the diagram at the left represents two consignments of fruit, each with four cartons. The empty squares represent un-infested units, and the squares with dots represent infested units. The four cartons in Consignment 1 have similar levels of infestation – this consignment is reasonably homogenous, and could be sampled as one lot. In Consignment 2, however, the two cartons at the right are much more infested than those on the left. This consignment is not homogenous. If it is sampled as one lot, the clumped infestation could be missed by chance; this consignment should be treated as four separate lots.

The NPPO specifying the sampling design does not know whether an individual consignment is homogenous in terms of infestation or not. The distribution of pests in lots is unknown, and so NPPOs may define a lot as being homogeneous in composition or origin[1], in an attempt to ensure homogeneity of infestation. These include commodity factors (species, variety, size, maturity), origin factors (country, place of production, grower, exporter, packing facility), and any treatments or processes applied.

The selection of criteria and degree of homogeneity required is up to the NPPO and will involve practical considerations, assumptions, knowledge of the commodity and packing procedures used for the lot being sampled. NPPOs should establish guidelines for their inspection staff on what they will accept as a homogeneous lot. For instance, an NPPO may assume that boxes of fruit in a shipping container from a single exporter but different packing facilities are homogenous for sampling purposes if the facilities operate a common quality system resulting in similar levels of residual infestation. This scenario might give an outcome similar to Consignment 1 in the diagram. Alternatively, the NPPO may know that the facilities are run by different companies with different systems, so that fruit from different facilities have different infestation levels (similar to Consignment 2). In this case, the NPPO may require that fruit from different facilities be sampled as different lots.

Finding pests in a sample generally results in the application of some phytosanitary action, such as treatment or reshipment, being applied to the entire lot. Conversely, if no pests are

[1] MAF Biosecurity New Zealand (2008) defines a lot as "The number of units of a single commodity (i.e. species), identifiable by such things as its homogeneity of composition and origin which forms part of a consignment."

found in a sample and other import requirements met, the entire lot is given clearance. If a non-homogeneous consignment, such as Consignment 2, is treated as one lot for sampling purposes, highly-infested parts of the lot may be given clearance if no pests are found in the sample. If pests are found, sections of the lot that have little or no infestation may have action taken unnecessarily. For this reason, it is important that a lot is defined in a way that provides some degree of homogeneity.

Sample Unit

Once the lot has been defined, the sample unit must be identified. A variety of factors affect the choice of the sample unit, some practical and others biological. For instance, if a lot consists of fruit in cartons, the unit could be either an individual piece of fruit or an entire carton.

If sampling is undertaken to verify that the consignment complies with phytosanitary requirements, then this should guide the selection of the sample unit. For instance, if the requirement is that less than a certain proportion of trays have infested fruit, the tray becomes the sample unit. If, however, the requirement is that less than a certain proportion of fruit are infested, the individual fruit is the sample unit. The form of the fruit will also influence the sample unit: grapes are typically shipped in bunches attached to a stem, rather than as individual grapes, so the sample unit might be a bunch or a box of bunches.

Whatever the choice of sample unit (box, bag, weight, bunch or individual fruit), the units must be independent and all have an equal chance of being selected for inspection if a statistical inference is to be made. This means that fruit packed in cartons in the centre of a container should have an equally likely chance of being selected for the sample as fruit in cartons packed near the container door. The next section discusses the implication of setting aside this requirement in favour of selecting units that are convenient.

Statistical and Non-Statistical Sampling

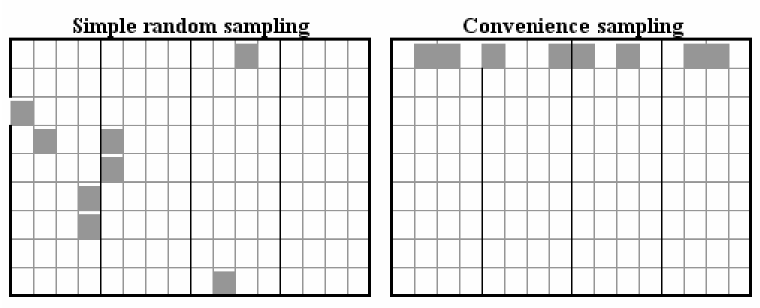
Typically, an NPPO will wish to make some inference about the condition of the lot based on the results of a sample. For instance, the NPPO may want a high degree of confidence that a pest, if present, does not exceed a particular level. To make this kind of statement, the sample must be representative of the lot in terms of infestation.

Defining the lot so that it is likely to be reasonably homogenous for pest distribution and using a statistical sampling method will help ensure selection of an unbiased sample that is representative of the lot. Non-statistical methods typically result in a biased sample, with the degree of bias unknown. Results of such samples can not be reliably extrapolated to make statements about the lot, unless the assumption is explicitly made that the degree of bias in the sample is negligible. While operational constraints may lead NPPOs to employ non-statistical elements in sampling, this reduces or eliminates the ability to draw statistical inferences about the condition of the lot from the results of the sample.

In general, sampling objectives to “provide confidence”, “verify compliance”, or “determine the level” suggest that a statistical sampling method should be used. If the objective is “to detect”, a non-statistical method may be sufficient – but with such methods, it is not possible to provide a quantified level of assurance, or confidence, about the level of pests in the lot.

Selecting a Sampling Method

A statistical sampling method is usually desirable for most quarantine situations. However, selecting independent sample units randomly may create operational difficulties. For operational reasons, samples are often taken from the first few boxes near the door of a shipping container, or from the tops of easily accessible boxes (e.g. convenience sampling), rather than unpacking the entire container to obtain a random and independent selection of fruit from different boxes.

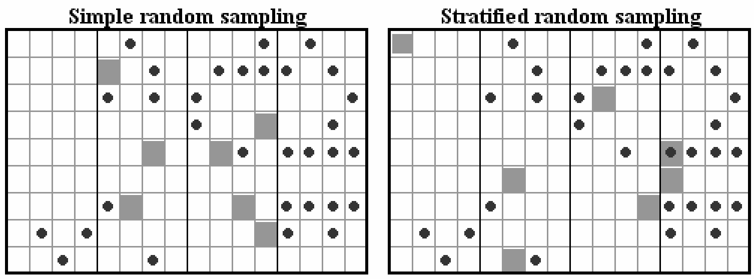


In the diagram at the left, the grey squares represent sampled units from two lots having four cartons each. Simple random sampling uses a random numbers generator (or other random process) to select the sample units from throughout the lot. This doesn’t ensure that sample units are evenly spread throughout the entire lot – by chance, multiple units may be selected from some cartons and no units from another carton, as in the diagram. An example of convenience sampling would be choosing all the sample units from the most accessible part of the cartons. This is operationally easier than simple random sampling, but if the consignment is not homogeneous with regards to infestation (for instance, if mobile pests have congregated in the lower layers of the cartons), this method is likely to produce a biased sample.

If a non-statistical approach is taken, the NPPO is essentially trading some of its assurance about the outcome of sampling for the speed, ease and/or lower cost of the convenient approach. It is not possible to determine the degree to which a non-statistical method reduces confidence in the result of any particular sample. Outbreaks of new pests associated with particular imports might serve as an indicator that sampling of such imports is not providing the required level of confidence.

In some cases, targeted (biased) sampling may be used to increase the chances of detecting a pest: for instance, deliberately sampling from areas of a lot that are considered more likely to be infested. However, targeting for one pest may reduce the chance of other pests being found. If nothing is found in a targeted sample, no firm statistical inferences can be made about the level of pests in the lot.

If a lot is not homogenous with regards to infestation, a simple random sample may (by chance) miss clumped infestation. An alternative statistical sampling method is stratified random sampling. The lot is divided into strata that could reasonably be expected to be more homogeneous internally with regards to pest distribution. Cartons or other packing units may serve operationally as different strata. A proportion of the sample units are then taken randomly from each stratum.



The diagram at the left shows the effect of two different sampling methods on a lot with four cartons. Black dots represent infested units in the cartons and grey squares indicate units selected for sampling. In each case seven sample units have been selected – however, with simple random sampling, all of those units may (by chance) miss one or more cartons. If the cartons have different levels of infestation, the results may not be representative of the entire lot. With the stratified approach, however, a certain proportion of the sample is taken from each of the four cartons. This reduces the likelihood that a carton with an unusually high infestation level compared to the rest will be missed by the sample. In practice, stratification may only be worth the effort for situations where infestation is likely to be infrequent and highly clumped, and the sample a very small proportion of the lot.

In selecting a sampling method, NPPOs should also take into account the likely distribution of pests in a lot. In many cases this will be unknown, but information about how the commodity has been treated in harvesting, sorting and packing may help guide the decision. For instance, fruit picked from an orchard straight into a bulk bin for export would vary in pest distribution as different areas of the field were picked. Clumps of infested fruit might be found throughout the bin, interspersed with clumps of un-infested fruit. However, if fruit from the same orchard is washed, sorted, graded and sized before being packed into cartons, the distribution of infested fruit might be considerably more random in the cartons than in the bulk bins. Simple random sampling might be sufficient for the fruit in cartons, while a stratified sample from the bin as it is being packed might be necessary to overcome the problems of clumping.

If there is no reason to suppose that infestation in a lot follows a set pattern (for instance, that every 10th fruit is infested), then systematic sampling is often easier than simple random or stratified random sampling, and produces a sample that is a reasonable substitute for a random sample. Systematic sampling involves dividing the number of units in the lot by the number of units in the sample to get the number of units per sample group (e.g., 30,000 units divided by 100 units gives 300 per group). An item is selected randomly from the first 300 units (say the 247th unit), and then every 300th unit is selected after that. In some cases, systematic sampling can be mechanically automated (e.g. every 50th fruit through a sorting machine is automatically diverted to a sample bin) for further ease and reduction of sampling bias.

For example, suppose a lot consists of 1000 boxes and the NPPO has determined that it will inspect 12 boxes (the sample unit in this case is the box). This means that approximately every 83rd box will be inspected (1000 ÷ 12 is approximately 83). The NPPO chooses a random starting point between 1 and 83, say box 25. The sample boxes are therefore box 25 and every 83rd box afterwards (e.g. boxes 25, 108, 191, 274, 357, 440, 523, 606, 689, 772, 855 and 938).

Sample Size Determination

To make a statistical inference with a particular degree of confidence about the infestation level of the lot, the size of the sample should be calculated based on specific pieces of information, known as parameters. These are discussed in the next section. Practical considerations, including cost, space and time, generally limit the number of units that can be sampled, particularly when units are destructively sampled (such as fruit being cut open). It may not always be possible to use inspection alone as complete risk mitigation due to the size of the sample required. In such cases, NPPOs should use additional measures, in combination with inspection, to achieve the desired outcome.

Sampling a fixed proportion of the lot (e.g. 2%), provides a consistent level of confidence in detecting a specific number of infested units, rather than a specific infestation level, and is discussed in more detail later.

Statistical parameters

Statistical sampling enables NPPOs to verify compliance with requirements about the infestation level of pests in imported lots with a particular degree of confidence. An example of such a requirement is given below.

At a 95% confidence level, not more than 0.5% of the units in the consignment are infested (MAF Biosecurity New Zealand, 2008).

Statistical sampling requires NPPOs to use a method giving an unbiased sample. The sample size is calculated from specific parameters, which are briefly introduced here, and then discussed in more detail in the following sections.

The two most commonly specified parameters are the desired confidence level and detection level (95% and 0.5% in the statement above, respectively). It is not enough simply to state that a sampling scheme should provide 95% confidence of detecting a pest. It must also include the infestation level at which the NPPO wants to be confident of detecting the pest.

The NPPO may also consider how many infested units would be tolerated in the entire lot, if any (the tolerance level), and the number of infested units that will be acceptable in the sample – the acceptance number. If the tolerance level is zero, the acceptance number will be zero. However, if the tolerance level is greater than zero, it follows that a number of infested units in the sample may be acceptable, depending on the sample size.

The NPPO must also consider how effective the detection method is. For instance, using a microscope to inspect leaves for mites might be highly effective, while using the unaided eye might be less effective. If the effectiveness of the detection method is not specifically used in the calculation of sample size, it is assumed to be 100%. This may not always be an appropriate assumption.

These parameters, the confidence level, detection level, detection effectiveness (and in some cases, the tolerance level and acceptance number), are used to calculate the sample size required for a statistical sampling scheme.

Confidence level

The confidence level is the probability that a pest infesting a specified proportion of units in the lot will be detected in the sample. A high level of confidence is typically desired, and 95% is commonly used. A 95% level of confidence indicates that, on average, if 100 samples are taken from a lot that has a specific proportion of units infested, 95 of the samples will detect the infestation, and 5 of the samples will not.

Higher levels of confidence may be appropriate in some circumstances, such as where the results of a risk assessment indicate a high degree of risk or uncertainty. The higher the confidence level, the larger the sample required to demonstrate it. The table at the right shows the effect of requiring 90%, 95%, 99% and 99.9% confidence of detecting an infestation level of 5% in a large lot.

Detecting a 5% infestation level in a large lot

Confidence	Sample size
90%	45
95%	59
99%	90
99.9%	135

The table illustrates the approximate relationship between sample size and confidence level in medium to large lots. A 99.9% level of confidence requires a sample three times as large as the sample required for 90% confidence, while a 99% level of confidence requires a sample size about twice as large.

A 100% confidence level is not possible, as sampling always involves error – there is always the possibility that an un-sampled unit is infested or that infestation is present but missed on a sampled unit, and so a value lower than 100% must be chosen.

Detection level

The detection level is the level of infestation upon which the sampling scheme is based. If the infestation level of the lot is equal to or larger than the detection level, the sampling scheme will detect at least one infested unit with the desired level of confidence. The detection level is expressed as a proportion or percentage of infested units in the lot, and must be greater than 0%. With sampling it is not possible to state, with any degree of statistical confidence, that 0% of units in the lot are infested. If no pests are found in a sample, the NPPO has the desired level of confidence that the infestation level in the lot does not exceed the detection level.

The choice of a detection level may be based on a tolerance level (see below) and is also often governed by practical considerations, such as resources available for sampling. A very low detection level (close to zero) requires a large number of units to be sampled to have a high probability of finding a pest, while a higher detection level requires fewer units to be sampled to detect it. The table below shows how the sample size changes with detection level in a large lot: in this example, a 95% confidence level is assumed in each case.

Sampling a large lot with 95% confidence

Detection level	Sample size		Detection level	Sample size
0.10%	2995		2%	149
0.50%	598		5%	59
1%	299		10%	29

A decrease in detection level requires an increase in the sample size to achieve the same confidence level. As a rough rule, halving the detection level requires twice the sample size.

The detection level may also be chosen to provide assurance about phytosanitary measures applied before inspection or testing. For instance, a treatment may be applied before shipment, and an NPPO may sample the consignment on arrival to determine that the level of live infestation is less than a particular level.

Tolerance level

The tolerance level is the degree of infestation in the lot that acts as the threshold for phytosanitary action, and can be expressed as either a number of infested units per lot, or as a proportion of infested units per lot. Some NPPOs have a zero tolerance for all quarantine pests, although sampling can never provide absolute certainty that the level of quarantine pests in the lot is actually zero. A tolerance level above zero means that pests present in the lot at levels below the threshold are considered acceptable by the NPPO. This situation could occur if a risk analysis has indicated that some pests have a very low chance of establishment, meaning that a low level of infestation is acceptable in imports. Tolerance levels may also be set for regulated non-quarantine pests or material such as soil, bark or other contaminants.

Although detection level and tolerance level are often treated similarly from a statistical standpoint, they are not the same concept. A tolerance level operates at the level of the entire lot, is based on an assessment of risk, and can equal zero. A detection level may be based on the tolerance level, but it is the level that the sampling scheme can detect, rather than what is considered acceptable. As explained above, the detection level cannot equal zero.

A non-zero tolerance level is related to an NPPO’s appropriate level of protection or acceptable level of risk, because the tolerance gives a quantified estimate of the number of infested units that the NPPO considers acceptable in imported commodities.

Efficacy of detection

The efficacy of detection refers to how effective the inspection or testing method is in finding infestation. It is expressed as the percentage of infested items inspected or tested that are correctly determined to be infested. While laboratory tests often have known levels of effectiveness (referred to as “sensitivity”), the same is not true for visual inspection. The issue of inspection effectiveness is mentioned in ISPM 23, *Guidelines for Inspection*, which states that the use of inspection is based on the recognition that there is some probability of pests being undetected. However, the assumption is frequently made that inspection will be virtually 100% effective, and detection efficacy is often overlooked in calculating sample sizes.

Detection of small or cryptic pests by eye can be difficult, particularly on commodities that have numerous hiding areas and may be difficult to inspect thoroughly. In such cases, an inspector may judge a sample unit to be un-infested by mistake. The efficacy of inspection methods is specific to the commodity and the detection practices used (such as time spent per unit, lighting, use of aids such as hand lenses or microscopes) and is unknown for most situations, although in some instances the effectiveness of detecting pests using specific techniques has been measured (for instance, see Gould, 1995). Reduced detection efficacy means that more units must be sampled to maintain the level of confidence in detection.

While the exact efficacy may not be known, it is worth making an assumption about inspection or test efficacy in situations where it is expected to be less than 95%, and increasing the sample size accordingly.

Acceptance Number

The acceptance number is the maximum number of infested units that will be accepted in the sample without the lot failing inspection. The acceptance number varies with tolerance level and sample size. In most sampling schemes, detection of a pest in a sample means the lot fails – the acceptance number is generally zero. However, in certain circumstances, an NPPO may decide to allow a certain number of infested units in the sample.

One such example is when pests of different status, or pests and contaminating material, may occur in the same lot. For high-risk pests, the tolerance level, and thus the acceptance number, may be zero, but a small quantity of low-risk pests or contaminant material, for which the tolerance level is greater, might be permitted in the sample.

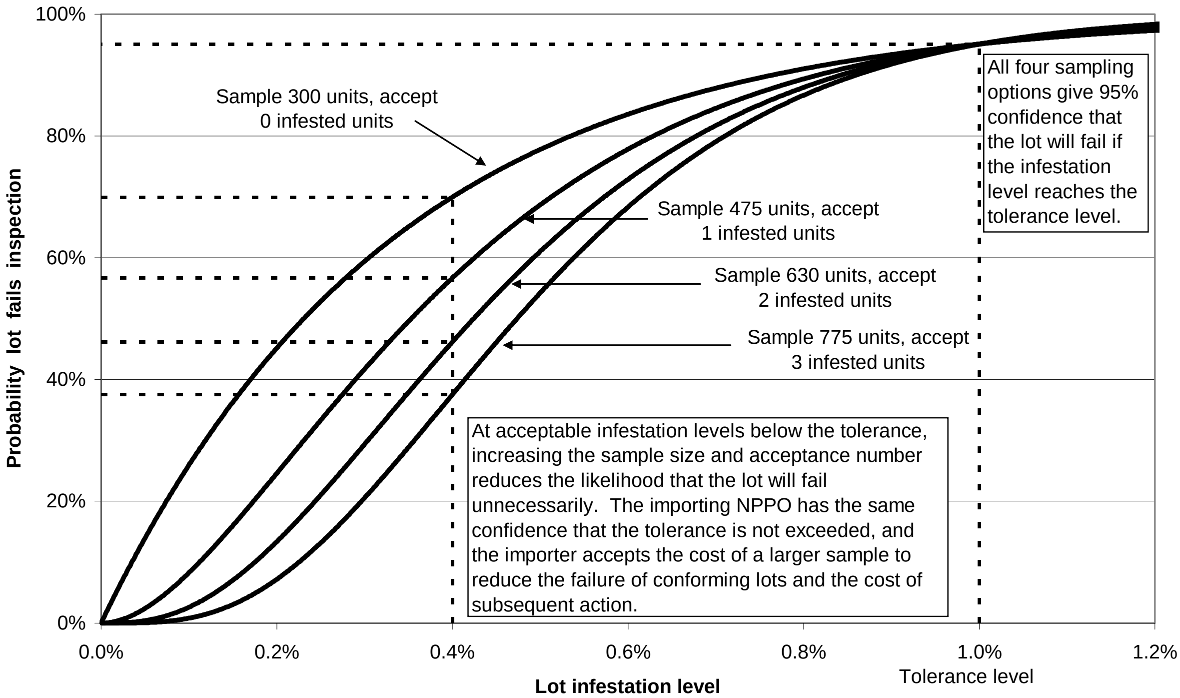
Another example of how acceptance numbers could be used in phytosanitary sampling is to provide greater assurance for exporters or exporting NPPOs that conforming lots are not rejected. An importing NPPO generally inspects the minimum sample necessary to provide a 95% level of confidence that a sample passing inspection does not exceed the detection level.

With a minimum sample size, finding a single pest or infested unit in the sample will cause rejection of the lot. If the desired tolerance level is zero, this is not an issue. However, if an importing NPPO has established a tolerance level for pests that is greater than zero, use of a minimum sample size will reject a number of lots that actually meet the tolerance. In such a situation, inspecting a larger sample and accepting some number of pests or infested units in the sample will reduce the chance of rejecting a lot that meets the tolerance.

For instance, a 600-unit sample will detect infestation of 0.5% with 95% confidence. If the tolerance level is 0.5% and the lot infestation level is only 0.3%, the lot meets the tolerance. However, a 600-unit sample still has an 84% chance of finding one or more infested units in the lot. This problem can be addressed with a larger sample size. As the sample size increases, the acceptance number also increases, providing a greater likelihood that a lot with infestation below the tolerance level actually passes inspection. While importing NPPOs would generally select the minimum sample to keep costs down, the importer (or the exporter) could find it cost-effective to pay for a larger sample to be inspected, reducing the likelihood of their shipments being rejected.

The chart below illustrates the impact of using larger sample sizes and acceptance numbers to reduce failure of conforming lots. The example uses a tolerance level of 1% and a confidence level of 95%.

Four different sampling schemes are shown: the minimum sample (300 units) has an acceptance number of 0. Larger sample sizes with acceptance numbers greater than zero still provide 95% confidence that infestation levels at the tolerance will be detected. However, at 0.4%, an infestation level well below the tolerance and therefore acceptable, the minimum sample size still has a 70% chance of failing the lot. However, sampling 775 units and accepting up to 3 infested units in the sample gives only a 38% chance of failing the lot.



From a statistical point of view, the importing NPPO has equal confidence that the tolerance of 1% has not been exceeded whether 0 infested units are found in a 300 unit sample or 3 infested units are found in a 775 unit sample. The importer may decide that reducing the failure rate for conforming lots justifies the cost of a larger sample.

Varying Level of Detection and Fixed Proportion Sampling

The detection level is usually a constant value for a particular import commodity (e.g. 1%); this means that the number of infested units per lot potentially entering the country will vary with the size of the lot and total quantity imported. If a particular sampling scheme results in lots passing inspection with an average of 0.1% of units infested, then importing 50,000 units results in an average of 50 (50,000 x 0.1%) infested units passing inspection. However, if imports double to 100,000 units, the number of infested units passing inspection also doubles, to 100 (100,000 x 0.1%).

With a constant detection level, the quantity of infested units being missed at inspection is dependent on the quantity imported. If risks are evaluated based on the number of pests arriving together at a given location and time, an NPPO may prefer to set the sample size to manage the number of infested units, rather than the percentage of the lot that is infested.

With a varying detection level, the importing NPPO would first determine the total number of infested units that would be allowable at a particular location in a particular time (such as a limit of 3 infested units per port per day). If 10,000 host units of a particular pest are imported at particular port in a day, then 3 infested units per day equates to a detection level of 0.0003. If only 1,000 units are imported at that port on the next day, then the detection level is 0.003. The sample size required to detect an infestation level of 0.0003 in 10,000 units (6,300) is 10 times as large as one required to detect an infestation level of 0.003 in 1,000 units (630). In each case, 63% of the lot is sampled.

Thus, while a fixed sample size enables one to make statistical inferences about the infestation level of a lot, sampling a fixed proportion of the lot enables inferences to be made about the number of infested units in the lot. However, this type of sampling should only be used where a specific number of infested units greater than zero are allowed to enter (e.g. a tolerance), as that number is required to determine the percentage to sample.

The tolerance used should be determined through risk analysis, and consideration should be given as to whether the number of infested units, or the percentage of imports infested, is more practical for risk management. Baker et al. (1990) give an example of a tolerance in terms of number of live fruit fly larvae per day; in this case, both inspection and treatment are required to ensure the tolerance is not exceeded.

Using the Tables

The appendices to the standard, while not an official part of the standard, are provided as examples. The associated tables can be used to find sample sizes for statistical sampling in order to provide assurance that the occurrence of pests in a lot is less than the level of detection. This is the common situation for which consignments are sampled at import or export. The tables are not exhaustive, and the appendices also provide methods of calculating other sample sizes based on specific values for confidence, level of detection and detection efficacy. For all values shown in the tables, the acceptance number is zero.

Tables 1 and 2 of Appendix 2 are used for relatively small lots (when the sample size is 5% of the lot size, or greater). For instance, suppose an NPPO wished to be 95% confident of detecting an infestation of 1% or greater in a lot of 1000 fruit, but had only 50% confidence in the effectiveness of the detection method for correctly identifying infested fruit. The number of units in the lot is 1000, the confidence level is 95%, and the level of detection multiplied by the efficacy of detection (1% x 50%) is 0.5%. These parameters are shown in Table 1 (see below), so the NPPO can determine the appropriate sample size from the table.

Table 1. Table of minimum sample sizes for 95% and 99% confidence levels at varying levels of detection according to lot size, hypergeometric distribution

Number of units in lot	P = 95% (confidence level) % level of detection x efficacy of detection					P = 99% (confidence level) % level of detection x efficacy of detection				
	5	2	1	0.5	0.1	5	2	1	0.5	0.1
25	24*	-	-	-	-	25*	-	-	-	-
50	39*	48	-	-	-	45*	50	-	-	-
100	45	78	95	-	-	59	90	99	-	-
200	51	105	155	190	-	73	136	180	198	-
300	54	117	189	285*	-	78	160	235	297*	-
400	55	124	211	311	-	81	174	273	360	-
500	56	129	225	388*	-	83	183	300	450*	-
600	56	132	235	379	-	84	190	321	470	-
700	57	134	243	442*	-	85	195	336	549*	-
800	57	136	249	421	-	85	199	349	546	-
900	57	137	254	474*	-	86	202	359	615*	-
1 000	57	138	258	450	950	86	204	368	601	990
2 000	58	143	277	517	1553	88	216	410	737	1800
3 000	58	145	284	542	1895	89	220	425	792	2353
4 000	58	146	288	556	2108	89	222	433	821	2735
5 000	59	147	290	564	2253	89	223	438	840	3009
6 000	59	147	291	569	2358	90	224	442	852	3214

Suppose instead that the lot was very large (300,000 units or more): the NPPO could use Table 1 again (below, left), choosing the value of 200,000+ for the lot size, which would give a value of 597 units for the sample size.

Table 1. Table of minimum sample sizes for 95% and 9[illegible] detection according to lot size, hypergeometric distribution

Number of units in lot	P = 95% (confidence level) % level of detection × efficacy of detection					P = % le dete
	5	2	1	0.5	0.1	5
25	24*	-	-	-	-	25*
50	39*	48	-	-	-	45*
10 000	59	148	294	581	2588	90
20 000	59	148	296	589	2781	90
30 000	59	148	297	592	2850	90
40 000	59	149	297	594	2885	90
50 000	59	149	298	595	2907	90
60 000	59	149	298	595	2921	90
70 000	59	149	298	596	2932	90
80 000	59	149	298	596	2939	90
90 000	59	149	298	596	2945	90
100 000	59	149	298	[illegible]	2950	90
200 000+	59	149	298	597	2972	90

Values in table 1 marked with an asterisk (*) have been rounde
resulting in a fraction of a unit being infested (for example, 3[illegible]
1.5 infested units in the shipment) are not possible. This means

Table 3: Table of sample sizes for 95% and 99% confi[illegible] according to efficacy values where lot size is large and suffi[illegible]

% efficacy	P = 95% (confidence level) % level of detection					P = %
	5	2	1	0.5	0.1	
100	59	149	299	598	2995	90
99	60	150	302	604	3025	91
95	62	157	314	630	3152	95
90	66	165	332	665	3328	101
85	69	175	351	704	3523	10[illegible]
80	74	186	373	748	3744	11[illegible]
75	79	199	[illegible]	798	3993	121
50	119	299	598	1197	5990	18[illegible]
25	239	598	1197	2396	11982	36[illegible]
10	598	1497	2995	5990	29956	91[illegible]

Alternatively, the NPPO could use Table 3 (above, right) or Table 4 of Appendix 3, which are based on the binomial and Poisson distributions, respectively, and select the value for 95% confidence, a 1% level of detection and 50% efficacy. The resulting values in this case are 598 (binomial) or 600 (Poisson). These distributions give virtually identical sample sizes to the hypergeometric distribution (Table 1) for large lot sizes.

POINTS OF DISCUSSION IN DEVELOPMENT OF ISPM NO. 31

Level of Detection and Tolerance Level

The question of whether the level of detection for a given sampling scheme also represented a *de facto* tolerance level generated considerable discussion. While these two concepts are both expressed in terms of the percentage of the lot that is infested with a pest, they differ from a regulatory perspective. Several members of the expert working group indicated that their NPPOs operated a zero tolerance for quarantine pests in imported consignments, even though the results of sampling could not provide full assurance that such a tolerance was met. In such cases, additional measures such as certification were used to increase assurance.

However, other NPPOs may use tolerances in particular situations, even for quarantine pests (for example, where the pests are present at a level too low to permit establishment, based on a risk assessment). For these NPPOs, the level of detection for sampling is set based on the tolerance for the lot, and the two terms may be considered interchangeable. In ISPM No. 31, the “level of detection” (what the sampling scheme can detect) and “tolerance level” (what the NPPO will accept) are presented as separate concepts, even though they may be, in some cases, the same number.

Acceptance Number

The use of acceptance numbers greater than zero is linked to the concept of tolerance level, and also generated considerable discussion. While most members of the expert working group indicated that their NPPOs would not use acceptance numbers greater than zero for quarantine pests, several used acceptance numbers for contaminating material such as weed seeds.

In this explanatory document, the acceptance number concept has been described in more detail for the benefit of NPPOs using, or trading with countries that use, a non-zero tolerance level. When sampling, officials generally want to sample the minimum number of units required to give a high degree of confidence that lots exceeding the tolerance level will be rejected. However, it is less well understood that the minimum required sample size also results in a significant chance that lots with an infestation level less than the tolerance will be rejected. As described earlier, accepting one or more pests in the sample and increasing the sample size appropriately provides the same confidence in rejecting non-conforming lots, while reducing the chance of rejecting conforming lots. Although importing NPPOs are generally only concerned to ensure that lots with infestation exceeding the tolerance level will be rejected, the importers, exporters and exporting NPPOs may be concerned about unnecessary rejection or treatment of conforming lots.

Aggregated Distributions

When pests are likely to occur in clumps, rather than randomly through the lot, their distribution is aggregated. Most statistical sampling methods are premised on a random distribution of pests. The more aggregated the distribution, the less reliable a method based on simple random sampling of the lot becomes. A stratified sampling plan can help reduce the impact of aggregation, but this relies on the NPPO being able to choose the strata in such a way as to minimise aggregation within strata. In general situations, ensuring that samples are taken from throughout the lot, rather than only the top of a bin or the boxes nearest the door, will help increase chances of detecting aggregated pests.

Appendix 4, while not a formal part of the standard, provides a calculation method for determining sample size if the degree of aggregation (between 0 and 100%) is known. While the degree of aggregation is very rarely known for general phytosanitary situations, this method may be useful for specific situations where pest aggregation information is available.

Fixed Proportion Sampling

Tables 5 and 6 in Appendix 5 are provided to inform NPPOs of the statistical implications of sampling a fixed percentage of the units in the lot (a commonly used method). This type of scheme gives varying levels of confidence for detecting a specific infestation level (Table 5) or different levels of detection for a given confidence level (Table 6) depending on lot size.

For instance, suppose an NPPO sets its sample size at 2% of the lot. If the lot has 1000 units, 20 will be sampled. Reading from Table 5, this sample provides 88% confidence that an infestation level of 10% would be detected (ignoring efficacy of detection). However, if the lot size was 100, the sample size would be 2, providing only 19% confidence that a 10% infestation level would be detected.

Alternatively, reading from Table 6, the sample size of 20 provides 95% confidence that an infestation level of 14% will be detected, while the sample size of 2 provides 95% confidence that an infestation level of 78% would be detected. As described earlier, fixed proportion sampling should only be used when the objective is to detect a specific number of infested units present in a lot, rather than detect a specific infestation level.

Hypergeometric Distribution Sample Sizes

The sample sizes in Tables 1 and 2, determined by the hypergeometric distribution, may appear counter-intuitive in some instances, because a lot of a given size may require a larger sample size than a slightly larger lot. The problem occurs because it is not possible to have a fraction of an infested unit, and so it is not possible to have exactly the same detection level for all lot sizes.

The hypergeometric distribution calculates the chance of detecting a given number of infested units in a sample of a given size, assuming that there are a specific number of infested units in a lot of a specific size. NPPOs generally set detection levels for sampling in terms of a percentage of the lot that is infested: this means that the number of infested units in the lot must be calculated.

For instance, if an NPPO desires 95% confidence of detecting pests infesting 0.5% of a lot, and the lot size is 800, this effectively means the sampling scheme should provide a 95% chance of detecting 1 or more infested units if there are 4 infested units (e.g. 0.5% x 800) in the entire lot. From Table 1, the required sample size is 421 units.

If a lot of 900 units arrives, the NPPO still wants 95% confidence of detecting pests infesting 0.5% of the units in the lot. In this case, however, 0.5% of the 900 units equates to 4.5 infested units – something that can not actually occur. The NPPO is therefore left with the choice of having 95% confidence of detecting an infested unit if there are 4 infested units in the lot (detection level 0.44%), or 5 infested units in the lot (detection level 0.55%). As the detection level of 0.55% associated with 5 infested units is more than the desired detection level of 0.5%, the NPPO would err on the side of caution and use the detection level of 0.44% associated with 4 infested units. A sample size of 474 gives 95% confidence of detecting one or more infested units when 4 infested units are present in a lot of 900.

However, when a lot of 1000 units arrives, the 0.5% detection level translates into 5 infested units. The sample size is now 450 – less than was required with a lot of 900 units. While this appears counter-intuitive, the smaller sample size is a result of the higher detection level. It takes fewer sample units to find 1 out of 5 infested units present in a lot of 1000 than to find 1 out of 4 infested units present in a lot of 900.

Lot size	Desired detection level	Infested units (calculated)	Actual detection level	Sample size
800	0.50%	4	0.50%	421
900	0.50%	4.5 (round down to 4)	0.44%	474
1000	0.50%	5	0.50%	450

Situations where this phenomenon occurs in Tables 1 and 2 of the standard are marked by asterisks.

REFERENCES AND ADDITIONAL INFORMATION ON SAMPLING CONSIGNMENTS FOR PHYTOSANITARY PURPOSES

Baker, R.T.; Cowley, J.M.; Harte, D.S. & Frampton, E.R. (1990). Development of a maximum pest limit for fruit flies (Diptera: Tephritidae) in produce imported into New Zealand. *Journal of Economic Entomology* 83: 13–17. This paper provides an example of how a non-zero tolerance can be translated into a required sample size, based on import quantity (related to the concept of a varying detection level). It includes consideration of treatment effectiveness, which is outside the scope of the sampling standard.

De Jong, P.D. (1995). Sampling for detection: leek rust as an example. *International Journal of Pest Management* 41: 31-35. This paper describes a simulation of different sampling schemes to detect a diseased plants in a clustered distribution in a field. In this example, a form of stratified random sampling provides better detection than simple random sampling.

Gould, W.P. 1995. Probability of detecting Caribbean fruit fly (Diptera: Tephritidae) infestations by fruit dissection. *Florida Entomologist* 78: 502-507. This paper gives an example of how the efficacy of a particular detection method can vary by commodity and by inspector.

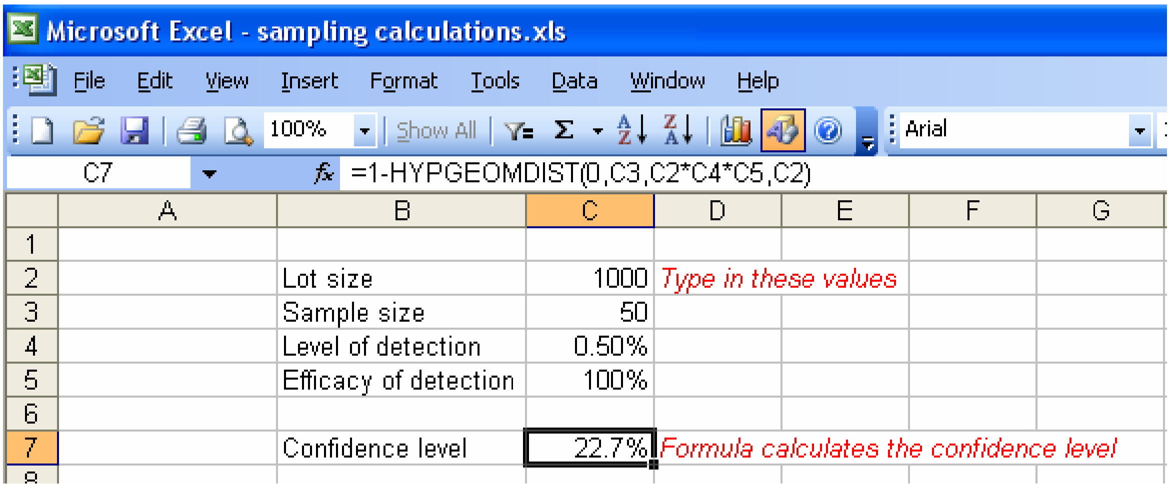
MAF Biosecurity New Zealand. (2008). Standard 152.02: Importation and clearance of fresh fruit and vegetables into New Zealand. http://www.biosecurity.govt.nz/files/ihs/152-02.pdf. This document provides an example of an NPPO's definition of a lot; an import sampling scheme based on a simplified hypergeometric table; the use of tolerances ("maximum allowable prevalence") for pests, soil and weed seeds; acceptance numbers; and different sampling options that can be requested by an importer.

United States Department of Agriculture. (2000). Sampling for the detection of biotech grains. http://archive.gipsa.usda.gov/biotech/sample2.htm This paper discusses sampling issues associated with detecting biotech kernels in grain lots, including concepts of tolerance, acceptance numbers and calculation of minimum sample size using a formula based on the binomial distribution.

Appendix. Using Microsoft Excel to Calculate Sample Sizes

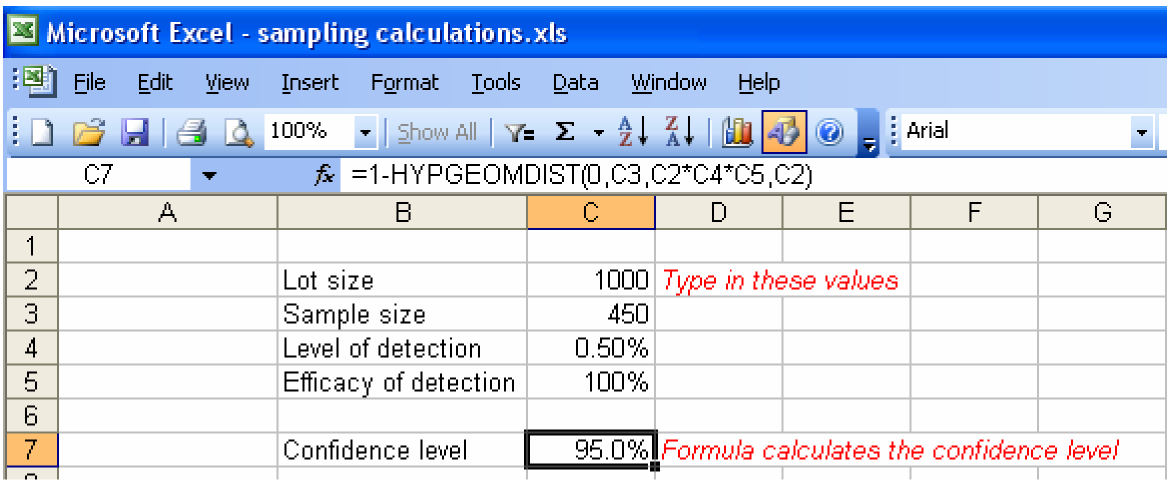
The sample size formulas given in Appendices 2 and 3 can be easily calculated using the statistical functions of Microsoft Excel.

<u>Hypergeometric Distribution</u>
The hypergeometric distribution should be used to calculate sample sizes for relatively small lots. The Excel function for the hypergeometric distribution calculates the probability of finding a specific number of infested units in the sample. If the number of infested units is set to zero, the function calculates the probability the lot will fail. One minus this probability is the confidence level.



Type in values for lot size, sample size, level of detection and efficacy of detection. In another cell, type in the formula, referencing the appropriate cells. The hypergeometric distribution calculation requires the number of infested items in the lot – this is calculated as the lot size multiplied by the proportion of the lot infested, or detection level, multiplied by the efficacy of detection. In this example, that quantity is C2*C4*C5, or 5.

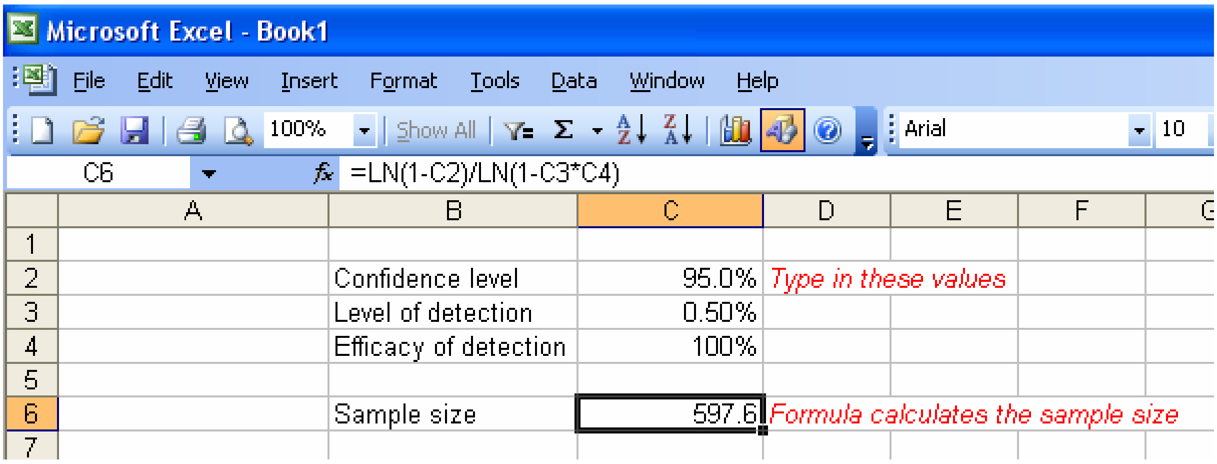
The confidence level is only 22.7%. To increase the confidence level, type in a larger value for the sample size. A sample size of 450 will increase the confidence level to 95%.



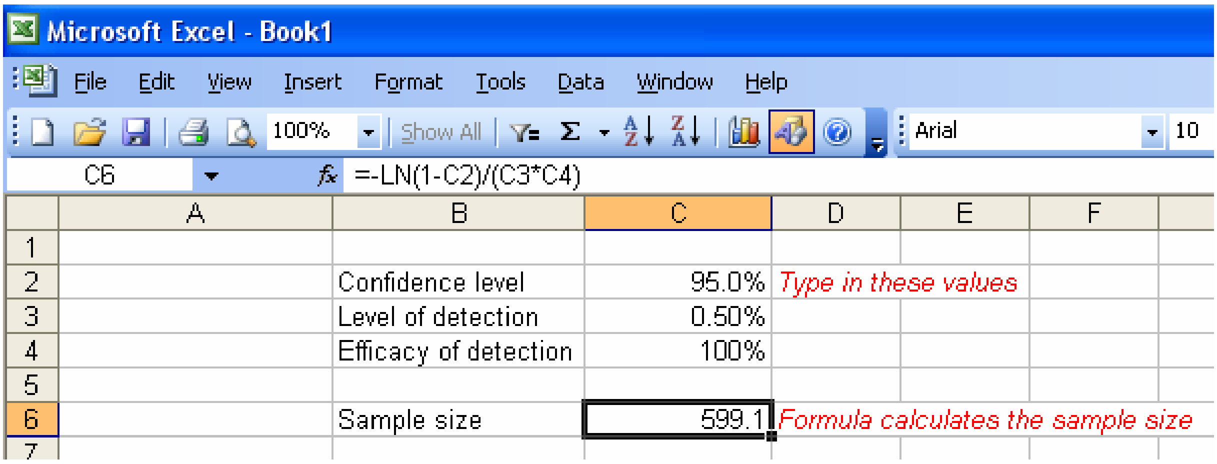
Binomial and Poisson Distributions

For larger lots, either the binomial or Poisson distribution can be used to calculate the sample size. It is possible to calculate the sample size directly if the acceptance number is zero, and to calculate the confidence level for both zero and non-zero acceptance numbers.

When the acceptance number is zero, the binomial sample size is calculated according to Formula 6 in the sampling standard, or as shown below in Excel. The sample size should be rounded up to the nearest whole number (e.g. 598 in the example below).

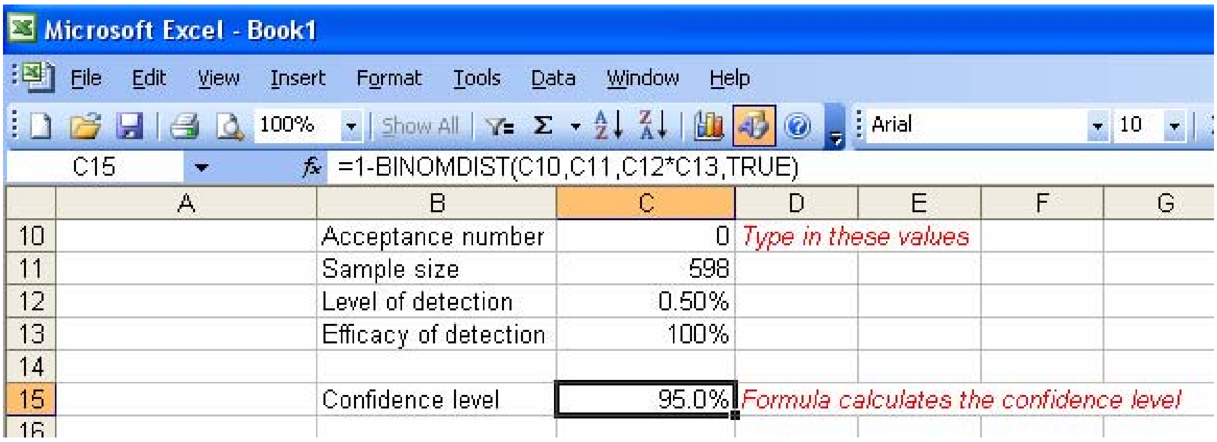


With a zero acceptance number, the Poisson sample size is calculated according to Formula 10 in the standard.

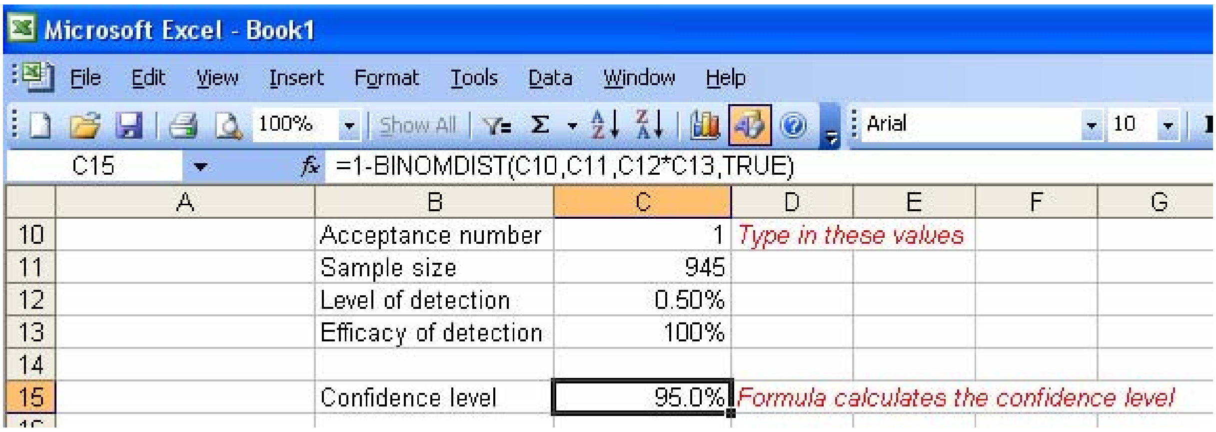


The binomial and Poisson distributions give similar sample sizes.

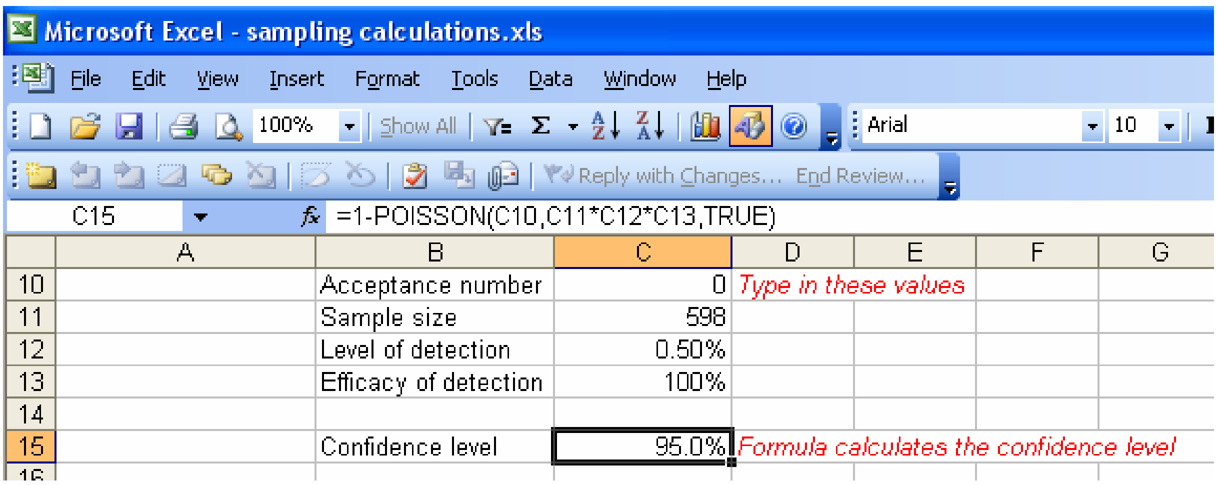
To include the acceptance number in the calculation, the confidence level is calculated, as for the hypergeometric distribution. Adjust the sample size number to obtain the desired confidence level.

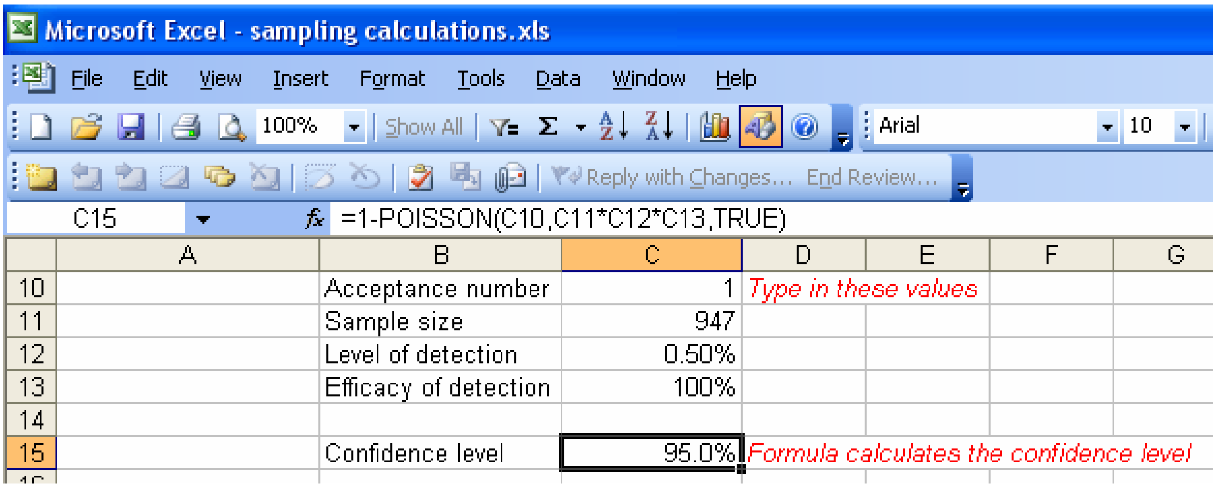


In this example, if the acceptance number is increased to 1, the sample size must be increased to 945 in order to maintain the confidence level at 95%.



The Poisson calculation formula gives similar results to the binomial distribution.




Microsoft Excel - sampling calculations.xls
File Edit View Insert Format Tools Data Window Help
100%
Show All
Arial
10
Reply with Changes... End Review...
C15
=1-POISSON(C10,C11*C12*C13,TRUE)
A B C D E F G
10 Acceptance number 1 Type in these values
11 Sample size 947
12 Level of detection 0.50%
13 Efficacy of detection 100%
14
15 Confidence level 95.0% Formula calculates the confidence level