Seed Science Research (2001) **11**, 101–119 DOI: 10.1079/SSR200166

PERSPECTIVES IN SEED TECHNOLOGY

Statistical considerations in seed purity testing for transgenic traits

Kirk M. Remund[1]*, Doris A. Dixon[1], Deanne L. Wright[2] and Larry R. Holden[3]

[1]Monsanto Company, 800 North Lindbergh Blvd., St. Louis, Missouri 63167 USA; [2]Pioneer Hi-Bred International, Inc., 7300 N.W. 62nd Avenue, Johnston, Iowa, 50131-1004 USA; [3]Sielken, Inc., Bryan, Texas, 77802 USA

Abstract

Recent advances in agricultural biotechnology have produced many new crop varieties with valuable transgenic traits. These varieties are being, and will continue to be, marketed alongside conventional non-transgenic varieties. As a result, seed purity in commercial seed lots is of particular importance to both seed consumers and seed producers. A key step in the seed production process is the design of sampling and testing procedures used to evaluate seed lot purity. However, due to uncertainties in such methods, there is always a risk of incorrectly rejecting or accepting a seed lot. This paper discusses factors that should be considered when designing and implementing seed purity testing procedures to manage this mis-classification risk – especially with regard to the presence or absence of transgenic traits. Many sources of uncertainty in both seed lot sampling and in the assay methods are described, and recommendations for reducing their impact are provided. This paper also explains the statistical concepts of misclassification risk as it affects seed producers and seed consumers. Sampling plans and formulas for determining the sample sizes necessary to control these misclassification errors when accepting or rejecting seed lots are also provided. Both simple, single-stage testing plans and the, often more efficient, double-stage testing plans are described. Testing seed pools rather than individual seeds is introduced as another way of adding efficiency to the testing process. Formulas are given for determining, from a seed sample, the confidence limits for the actual purity level of a commercial lot.

Keywords: acceptance sampling, assay error, biotechnology, GMO seed purity analysis, seed technology, seed testing, transgenic seeds

*Correspondence:
Fax: +1 314–693–6673
Email: Kirk.M.Remund@monsanto.com

Introduction

A key step in the seed production process is to determine if a seed lot meets minimum purity requirements with respect to various criteria. In the past, such properties as percentage germination, noxious weeds and inert matter have been evaluated. In the present and future, more seed varieties containing transgenic traits, such as Roundup Ready Soybeans®, will be marketed alongside conventional seed varieties. Because of this development, genetic and conventional seed purity, as it relates to the presence and absence of transgenic traits, is being added to the list of properties to be measured. This paper focuses largely on testing to confirm the absence of transgenic traits in conventional lots; however, the methods described can be adapted for other applications, including testing for the presence of conventional or off-type transgenic seeds in a transgenic seed lot.

Seed testing labs have used tolerance tables, such as those published by the Association of Official Seed Analysts (AOSA, 1998) and International Seed Testing Association (ISTA, 1999), to compare estimated seed lot purity levels to labelled purity claims (e.g. 90% germination claim). The methods presented in this paper generalize this AOSA and ISTA approach to include incorporation of assay errors, seed pooling, risk and cost management, and multiple-stage testing. The methods in this paper can be used to satisfy the very high expectations for genetic trait purity in transgenic and conventional seed varieties testing.

This paper discusses testing methods that can be used to estimate genetic purity of a seed lot as well as to set accept/reject criteria based on a given threshold. Statistical considerations that are needed to design and carry out reasonable sampling and testing plans are also discussed. While the focus is on transgenic trait purity, these methods also have application to purity testing for other types of traits. It

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

is not the purpose of this paper to provide testing plan recommendations for specific seed testing applications, but examples for illustrative purposes; nor are specific economic or business issues associated with transgenic seed testing discussed.

This paper is organized as follows. The first section presents the general concept of seed lot acceptance sampling; this section also discusses seed lot acceptance decisions in the presence of sampling uncertainty and the associated risks of making inappropriate decisions. The second section introduces protein and DNA assay methods used to measure the presence of transgenic traits and statistical implications of using each of these assays. This is followed, in the third section, by a discussion of different types of testing plans and the impact of assay system error rates on these testing plans. The fourth section contains a discussion of seed lot testing plans designed for classifying a lot (i.e. accept or reject the lot) versus estimating the purity level of the lot. The key sampling and testing plan assumptions are enumerated in the fifth section, which also contains a spreadsheet application example illustrating the use of these methods in designing seed testing plans. Finally, the Appendix includes the statistical formulas used in the calculations for seed purity statistical tests.

Seed lot testing uncertainty and the associated risks

Figure 1 is a simple illustration of a seed lot sampling and testing plan. A sample of seeds is taken from a lot. The seeds are tested individually or in seed pools using an appropriate protein or DNA assay. If the number of deviant seeds or seed pools (e.g. those containing an undesirable genetic trait) exceeds a cut-off point (c), then the lot is rejected; otherwise, the lot is accepted. The decision to accept or reject the seed lot is based entirely on results obtained from assays of traits in the seed sample. If the purity in the sample is not representative of the lot bulk characteristics, or the assay system has a high error rate that is left unaccounted for, then the lot will have a greater than expected chance of being misclassified. Even under ideal conditions, there will always be risks of accepting impure lots or rejecting adequately pure lots. The only way to avoid these risks entirely is to test all of the seeds in the lot with an error free and non-destructive assay system. This is rarely possible and almost never practical. Since these risks cannot be eliminated, the alternative goal is to maintain the risks at acceptable levels. This can be accomplished in part by choosing an appropriate sampling and testing plan.

Three primary contributions to the risks mentioned above are given by random sampling effects, non-random seed lot sampling effects and assay system uncertainty. Seed testing plans discussed here are based on the assumption that the seeds tested are a simple random sample from the entire seed lot. Uncertainty resulting from testing such a random sample will always be present, but can be modelled well with a binomial probability distribution; this distribution is incorporated into the statistical testing procedures described in this paper. When the seeds tested are not truly from a random sample, the risk of making an incorrect decision is increased due to non-random sampling. Finally, assay system uncertainty affects the ability to correctly classify a seed (or seed pool) as being pure or impure; when assay system error rates are known and small, they can be effectively accounted for in a statistical testing procedure.

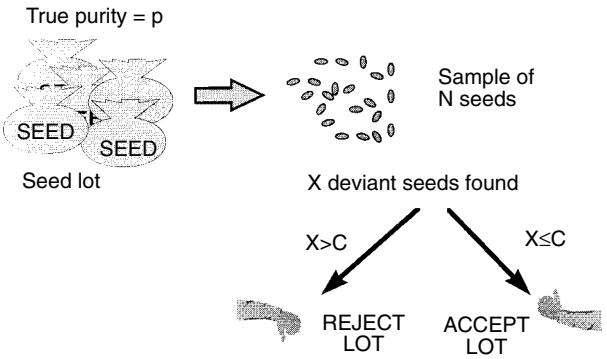


Figure 1. Basic seed lot sampling plan. If the number of deviant seeds exceeds c, then the lot is rejected.

The objective of this section is to help the reader gain an appreciation for these sources of uncertainty and their associated risks. A second objective is to provide direction to minimize these sources of uncertainty in the planning stages of seed lot sampling schemes.

For the purpose of discussing these contributions to risk, the distinction between seed 'sampling' and 'testing' plans is described here. In the context of this paper, the seed sampling plan is the physical procedure that is followed to take seed samples from a lot. The number of seeds tested, reject/accept criteria, sample preparation steps and method used for testing are all elements of the testing plan.

Random sampling uncertainty

This section discusses the management of random sampling uncertainty, which is inherent when sampling any seed lot, by incorporating this uncertainty into appropriate statistical testing procedures. A key objective is to design a testing plan

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

that has satisfactory risk levels in the presence of both random sampling and assay system uncertainty. Here the focus is on the management of random sampling uncertainty. Assay system uncertainty, and how to incorporate such uncertainty into a testing plan, are addressed in later sections of this paper.

A basic seed lot testing plan has two key parameters: (1) the number of individual seeds (or seed pools) to sample and test; and (2) the maximum number of deviant or unacceptable seeds or seed pools that can be tolerated in the sample before the seed lot is rejected. For example, one may choose to test 400 individual seeds from a lot and reject the lot if more than four deviant seeds are observed. How good is this testing plan? This question cannot be answered until there is a definition of what is meant by 'good' and 'bad'. The following definitions, based on standard acceptance sampling terminology in statistical quality control literature (e.g. Montgomery, 1997) are needed to accomplish this:

- Lower quality limit (LQL): the lowest level of purity in the seed lot that is considered acceptable to the consumer. This could be the minimum level of purity that needs to be defended statistically, legally, etc. If the seed lot is accepted, the lot can be labelled as having a level of purity at or above this LQL with a high degree of confidence. The LQL is often called the purity (or impurity) threshold.
- Acceptable quality level (AQL): the lowest level of purity in a seed lot that current production practices can support. Seed producers want to have a high probability of accepting lots with purity levels at this level or greater.
- Producer's risk: the chance of rejecting a seed lot that has actual purity at the AQL. Simply stated, this is the probability of rejecting a nearly 'pure' seed lot.
- Consumer's risk: the chance of accepting a seed lot that has actual purity at the LQL. This is simply the probability of accepting an 'impure' seed lot.

In practice, appropriate values for the LQL and AQL are not always obvious. Ideally, the consumer may prefer complete purity (i.e. LQL = AQL = 100%), a goal that will generally not be achievable due to practical limitations. Instead, the LQL should be set at the lowest level of purity that the consumer is willing to accept; government regulatory agencies may provide guidelines about appropriate LQL levels. The AQL should be set at a level that current production practices can support and may be based on estimated impurity levels in previous lots. To design a reasonable testing plan in terms of both consumer's and producer's risks, the AQL purity level must be larger than the LQL purity level; otherwise it will be difficult, if not impossible, to produce seed that is pure enough to meet the consumer's expectations.

It is important to design a testing plan that minimizes both the consumer's risk and the producer's risk at practical levels of the AQL and LQL. For example, suppose the AQL and LQL are set at 99.5% and 99.0%, respectively (or stated analogously in terms of impurity, at 0.5% and 1.0%, respectively). One may require a testing plan to reject at least 95% of the seed lots with purity levels at or below the 99% LQL (i.e. 5% consumer's risk) and accept at least 90% of the seed lots with purity of 99.5% or greater (i.e. 10% producer's risk). This is one example of criteria that are used to establish testing plans; in practice the criteria will vary given the different objectives that exist for seed lot acceptance sampling. As illustrated here, the AQL and LQL can be discussed both in terms of purity and impurity levels. Statistical quality control literature tends to focus on the percent defects or, in this case, impurity. In this paper, examples using both metrics will be provided.

A graphical tool used to evaluate testing plans against established criteria is called an operating characteristic (OC) curve. Figure 2 shows an example of an OC curve for one possible testing plan (solid curve). The true impurity in a seed lot is plotted on the horizontal axis, while the probability of accepting the seed lot is plotted on the vertical axis. The solid curve then shows the probability that a proposed testing plan will result in acceptance of a lot at each specified level of lot impurity. The testing plan associated with the curve in Fig. 2 is desirable. It has a high probability of accepting a lot that has impurity at the AQL (i.e. a producer's risk of only 5%) and a low probability of accepting a lot at the LQL (a consumer's risk of only 5%). The ideal OC curve (thick dashed line), which would be achieved only if the entire lot of seed was tested, is also plotted as a comparison to the proposed testing plan curve. This ideal OC curve gives a 0% chance of rejecting a lot with true impurity less than the LQL and a 100% chance of rejecting a lot with true impurity greater than the LQL.

Acceptance sampling concepts are further introduced here using examples. In these examples, an assumption is made that the assay system can distinguish perfectly between deviant and non-deviant seeds (i.e. assay system error rates are zero). A later section will discuss how non-zero assay error rates can be incorporated into a testing plan. Producer's and consumer's risk probabilities are calculated based on binomial distribution probabilities, and the formulas used in the calculations are provided in the Appendix. The binomial distribution provides very reasonable estimates of these probabilities as long as the sample taken from a lot is no larger than 10% of the lot population (Montgomery, 1997). Therefore, the statistical methods described in

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

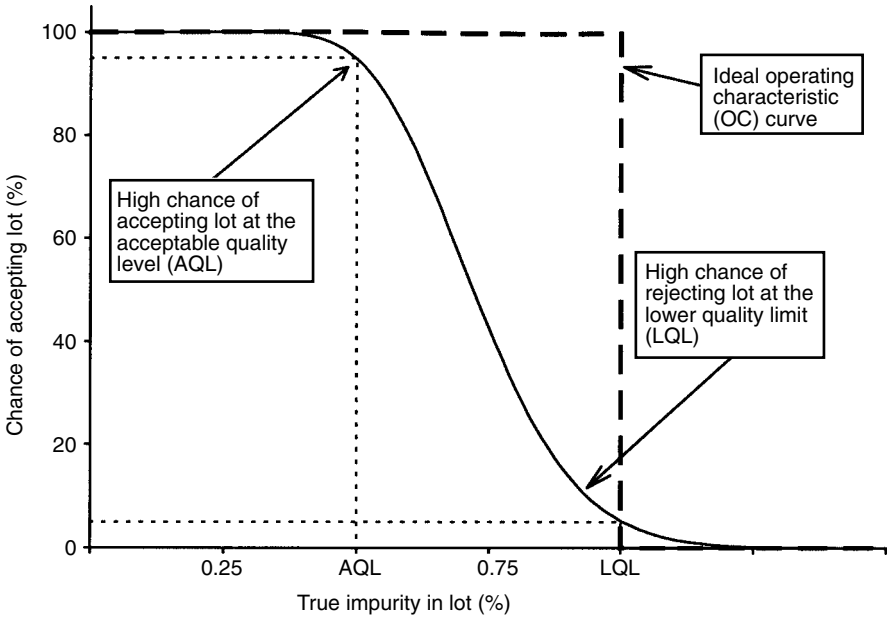


Figure 2. Example of a testing plan operating characteristic curve with low producer's and consumer's risks.

this paper can be used regardless of how small the seed lot size is, as long as this condition is met. If this condition is not met, analogous (albeit tedious) calculations, using the hypergeometric distribution, can be substituted for the binomial (Montgomery, 1997).

Suppose that a proposed testing plan is to sample and test 400 individual seeds from a lot and reject the lot if more than 1% of the seeds (i.e. more than four seeds) in the sample are deviant. The AQL and LQL are set at 0.5% and 1% impurity (or 99.5% and 99% purity), respectively. The OC curve for the testing plan in this example is plotted as a dashed curve in Fig. 3. This curve shows that this testing plan would have a consumer's risk of 63%; that is, seed lots with 1% impurity will be accepted 63% of the time. Even lots with 1.5% impurity will be wrongly accepted 28% of the time. On the other hand, this testing plan has a producer's risk of 5%; that is, seed lots with 0.5% impurity will be rejected only 5% of the time. While this testing plan is reasonable in terms of producer's risk, it is obviously very poor in terms of the consumer's risk. If this testing plan is modified to reject seed lots that have more than 1 deviant seed in the sample, then a curve is obtained that has a relatively low consumer's risk but a high producer's risk (60% producer's risk; see Fig. 3, thin solid curve). By increasing the number of seeds tested to 3000 individual seeds and the number of permitted deviants to 21, a curve can be obtained with both a low producer's and a low consumer's risk (Fig. 3, thick solid curve). Table 1 demonstrates how the producer's and consumer's risks are reduced as the number of seeds tested and number of deviant seeds permitted are increased with an LQL of 1% impurity and an AQL of 0.5% impurity.

In many seed testing situations, the number of tested seeds required to minimize producer's and consumer's risks may be impractical given the available testing resources. For example, testing 3000 individual seeds is required in the previous example to achieve 5% consumer's risk and 5% producer's risk (Fig. 3 and Table 1). This may be practical in some testing situations, for example when examining seeds visually for noxious weed seed contamination. However, when expensive genetic assays are required, the cost of testing 3000 individual seeds per lot is not currently a practical option. This paper describes how seed pool testing may reduce the amount of seed testing resources needed, while maintaining sufficiently low error rates.

It is sometimes advocated that all genetic purity test plans should result in rejecting lots when a single deviant seed is found in the sample (i.e. $c = 0$). Pressure to use such plans stems from the belief that detecting even a single deviant in a sample tells the consumer that the lot is impure. One should not, it is argued, accept a known impure lot. A few words of caution regarding the implementation of such zero-tolerance testing plans are necessary. First, these testing plans are usually associated with very high producer's risks. Figure 4 illustrates this point for a

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

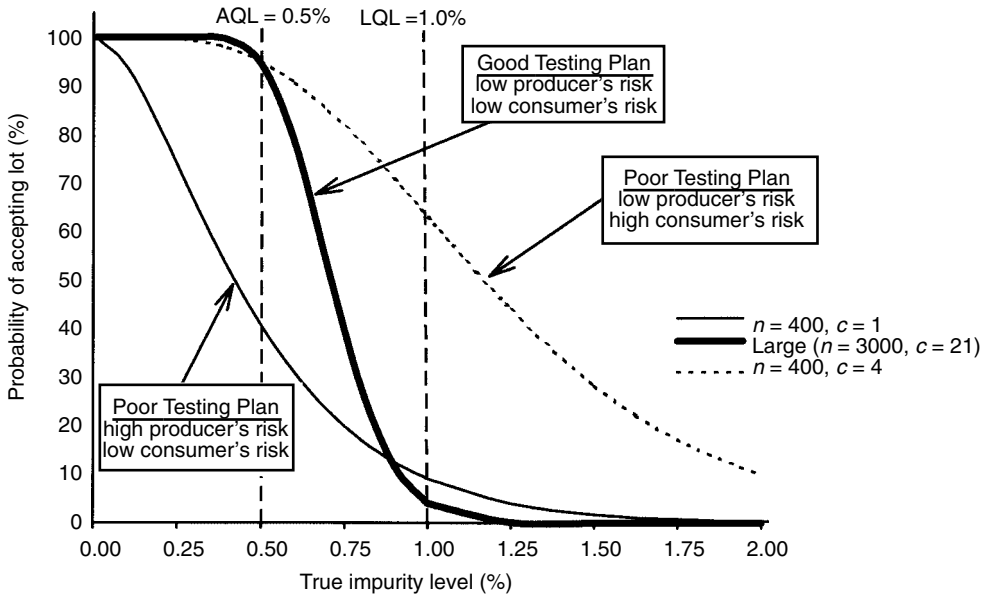


Figure 3. Operating characteristic curves for three different seed lot testing plans with varying producer's and consumer's risks.

Table 1. Consumer's and producer's risks as the number of tested seeds increase. The number of deviant seeds for the rejection criterion is kept roughly proportional to the number of seeds tested

Number of seeds tested	Reject lot if deviants exceed (*c*)	Producer's risk (%)	Consumer's risk (%)
200	1	26	40
400	2	32	24
800	5	21	19
1600	11	11	13
3000	21	5	5

testing plan designed to classify a seed lot at a 1% LQL impurity threshold. In this example, 400 individual seeds are to be tested at an AQL of 0.5% and a LQL of 1.0%, with no allowable deviants accepted. The consumer's risk for this zero deviant testing plan is low (i.e. seed lots with 1% impurity will be accepted 2% of the time). However, the producer's risk for this testing plan is extremely high (i.e. 87% of the seed lots at 0.5% impurity would be rejected). The AQL would need to be changed from 0.5% to 0.01% impurity in order to reduce the producer's risk to 4%. In other words, seed production practices would have to be in place to achieve seed purity in the range of 99.99% or higher, a level that would be extremely difficult (perhaps impossible) and very expensive to achieve for many crops. There may be occasions where zero deviant testing plans may be appropriate; however, the associated producer's risk should be considered along with the consumer's risk to determine if such a testing plan is reasonable. Secondly, there is a misconception that if there are zero deviants found in the test results, then there are zero deviants in the whole lot under consideration. The only way to conclude that there are zero deviants in a seed lot is if every single seed in the entire lot is tested as a non-deviant. This would require a non-destructive test that is practical to implement on every seed in the lot.

Non-random sampling uncertainty

The Association of Official Seed Analysts (AOSA, 1998) indicates that no matter how accurate a seed testing procedure may be, it is only as good as the

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

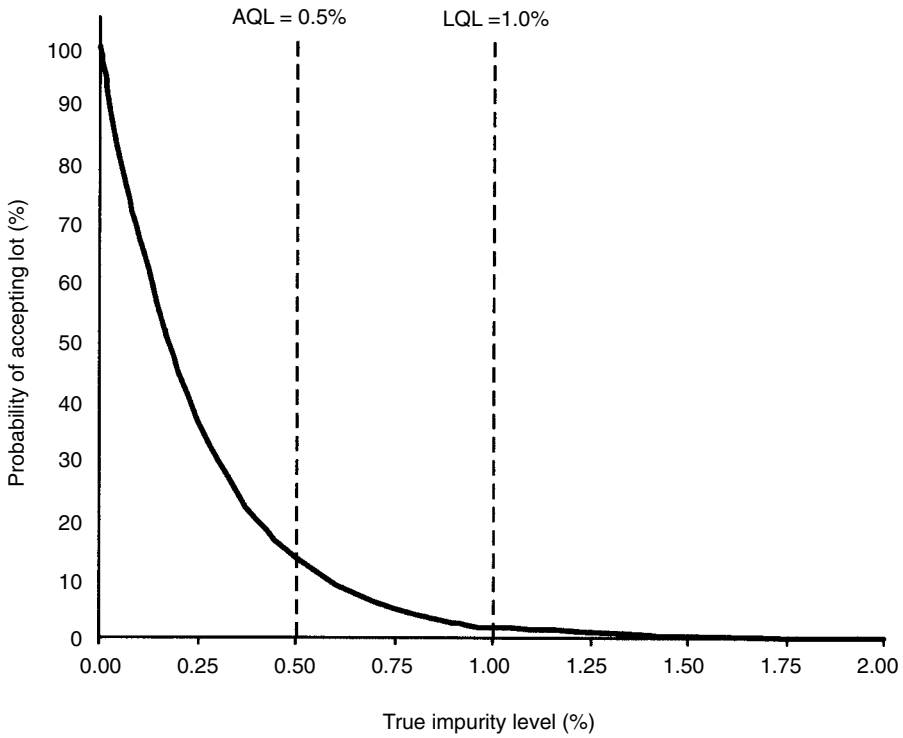


Figure 4. Operating characteristic curve for a zero-deviant testing plan. The curve shows that this type of testing plan generally has a high producer's risk (acceptable quality level = 0.5% and lower quality level = 1.0%).

sample taken from the lot. For this reason, it is critical that seed samples are taken that collectively represent bulk seed lot characteristics. Unfortunately, 'representative' is a somewhat vague and non-statistical term. This section describes sampling methods that can be used to obtain a 'representative' sample in a statistical sense.

A sample that has the same proportion of deviants as the lot would certainly be representative of the lot. However, in general, the percentage of deviant seeds in a sample will differ from the actual lot impurity due to random sampling variability. In order to correctly determine the values of consumer's and producer's risk discussed in the previous section, there is a need to describe the variability associated with a given number of deviant seeds in a sample in addition to how close the sample 'impurity' is expected to be to the lot impurity. The previous section discusses statistical methods for incorporating this random sampling variability into sampling and testing plans. This section describes methods to obtain a sample that has 'expected' impurity equal to the impurity in the lot, even though the 'actual' impurity of the sample will likely not be equal to the lot impurity.

A purely representative sample is defined as one that was selected using the method of simple random sampling from the bulk seed lot. Simple random sampling implies that each seed in the lot has both an equal and an independent chance of being included in the sample. Consider the following example. Suppose a single scoop of 2000 seeds is taken from some random position in the seed lot. Such a sample would be a random sample since every seed in the lot has the same chance of winding up in the scoop; however, it is not a simple random sample since the seeds are not independently chosen into the sample. A seed from the north end of the lot could never be in the same sample as one from the south end. If deviant seeds tended to be found close together in the lot, such a sample would either have no deviant seeds or a very large proportion of deviant seeds. Such large sample-to-sample variation (i.e. 0% or 100% sample purity) would be much different than that described by the binomial distribution under simple random sampling.

The statistical methods used in seed testing (see the Appendix) and the associated inferences are based on the assumption of simple random sampling from a seed lot. If such sampling were possible, the binomial probability model would always be valid even if the seed lot was heterogeneous. To obtain a simple random sample of seeds, however, each individual seed in the lot might need to be indexed and

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

accessible for independent random selection. This is obviously impractical for any normal-sized seed lot. The extent to which the sampling methodology deviates from simple random sampling and invalidates the probability models used is precisely the effect that is called non-random sampling error. Sampling methods that minimize the effect of non-random sampling will now be given.

Given that simple random samples are impractical, there are practical ways to reduce non-random sampling error. For example, in the previous 'one-scoop' example, if the seed lot were thoroughly mixed prior to sampling, then the lot would be completely homogeneous and the scoop would contain a simple random sample. Unfortunately, nature does not generally provide homogeneous seed lots, and mixing large lots is not always easy or practical. It might, however, be reasonable to obtain lots that have moderate degrees of homogeneity. A reduction in the heterogeneity of the bulk seed lot will reduce the impact of non-random sampling. It should be noted that highly heterogeneous lots are problematic for other reasons as well, and a single purity value should be interpreted with caution if the lot is not known to be homogeneous.

In general, sample selection methods can be used that mimic the characteristics of a simple random sample as closely as possible, although the particular method used will depend on when the seed samples are taken during the production process. For example, a seed sample would be collected differently if seeds were obtained in the field versus in a railway wagon versus in a grain silo. The sample selection plan is chosen based on the spatial arrangement of the seeds and their accessibility. Two sample selection methods that will cover many seed testing situations are discussed next.

Probe sampling can be used if seeds are in a large open container such as a truck or a railway wagon. The probe sampling scheme is somewhat related to stratified and cluster sampling schemes (Cochran, 1977) which are commonly used in statistical sampling plans. The surface of the seed container is divided into an invisible grid of squares, and a probe is used to take samples within each grid square. The probe will recover seeds from several depths in the seed bulk. It is important to note that the number of sampling points in a grid should be chosen so that all phases anticipated in the seed lot may be represented in the sample. In stratified sampling, the seed lot is divided into subunits called strata. This subunit division should be made along boundaries that separate known phases in the seed lot. In practice it is difficult to know where these phase boundaries are located, or if they even exist at all. For this reason a probe sampling plan is advised.

A systematic sampling scheme can be used to sample large 'closed' containers such as a cargo hold on a ship or a silo. This type of a sampling container does not fit easily into a grid format. However, a continuous seed flow that is generated as the container is filled or emptied can be sampled. For example, the continuous seed flow may be sampled every 5 minutes as seeds are removed from a silo. A continuous dribble sample may also be taken from the continuous flow. As with the probe sampling scheme, the total number of samples taken should be dependent on the number of expected phases in the seed lot. If there are several phases in the seeds (i.e. much spatial heterogeneity), then more samples should be taken than in the case where the seed lot is judged to have few phase changes (homogeneous seeds). To obtain more detailed instructions on sampling techniques, see United States Department of Agriculture (USDA, Grain Inspection, Packers and Stockyards Administration, 1995a, b) and International Seed Testing Association (ISTA, 1999) publications.

Usually the goal of seed purity testing is to assess the overall quality of seed lot bulk characteristics. In this situation, seed samples collected (regardless of the sampling scheme) can be mixed together in a composite sample. This composite sample should be thoroughly mixed. Individual seeds or seed pools are taken from this composite sample for testing using an appropriate assay method.

Finally, if the goal of the seed purity testing is to assess how evenly distributed the seed quality is across the lot, or to test for areas with higher levels of impurity or uneven distribution of a particular trait in the seed lot, then mixing the seed samples is inappropriate. Methods for such within-lot homogeneity analyses are not discussed in this paper.

Assay system uncertainty

The previous two sections were concerned with seed testing uncertainty due to random and non-random sampling effects. Seed testing uncertainty is also affected by errors that occur during the process of assaying seeds for the desired purity attribute. The assaying system includes all of the sample preparation steps as well as the physical testing of seeds using laboratory instruments or bioassay methods. This section will discuss the uncertainty associated with assay systems.

There are usually multiple steps taken to prepare seed samples for testing. At many of these steps, there is the possibility of mislabelling, cross-contaminating or mishandling the sample. This is particularly true for testing pools of seeds rather than individuals. Many of the testing methods for DNA and protein require that the seed samples be ground into flour. In the case of large seed pool samples, the seeds must be ground appropriately to provide a flour particle size

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

that consistently generates a homogeneous mixture for seed testing to avoid introducing assay system errors. There is also the possibility that dust from flour in a sample may mix with the seed or flour from another sample and contribute to assay system errors.

The laboratory method that is used to test prepared seed samples will also introduce some level of uncertainty to the test results. The laboratory method has limits to its sensitivity and specificity, and steps should be taken to understand the limits of the testing method. Again, this is particularly important for large seed pool samples where the impurity characteristic could be found at trace levels.

All of these sources of assay system uncertainty are revealed in one of the following two types of error rates:

- False-positive rate: this is the likelihood of a seed or seed pool testing positive for an impurity characteristic when in reality it is negative. One hundred percent minus the false-positive rate is often called the specificity of the assay.
- False-negative rate: this is the likelihood of a seed or seed pool testing negative for an impurity characteristic when in reality it is positive. One hundred percent minus the false-negative rate is often called sensitivity of the assay.

All of the steps in an assay system have an effect on both of these error rates. In many cases, the false-positive rate can be dealt with effectively by simply re-testing an independent aliquot of the sample to confirm a positive result. For example, if the false-positive rate for the assay system was 5% (quite high), an automatic retest of positive results from an independent aliquot of the sample could reduce this false-positive error rate to 0.25% (i.e. $0.05 \times 0.05 = 0.0025$). In the context of testing for transgenic traits, this large reduction in the false-positive rate may not be realized if adequate measures are not taken to clean seed grinders thoroughly between seed samples, since residual traces of transgenic traits may remain in the grinder from previous positive samples.

Evaluation of the false-positive and false-negative rates should be carried out before seed lot testing begins. Several factors can influence assay error rates. Different seed pool sizes need to be evaluated to determine how large the pool size can be and still have an acceptably low false-negative rate for a given assay. The false-positive rate and the false-negative rate will both contribute to the calculation of the producer's and consumer's risks, so it is very important that they are negligibly low (e.g. 1%). Testing plans can be designed to account for negligibly low known error rates. While such adjustments will allow valid sampling plans to be developed, they will not remove the loss in statistical power due to assay errors (i.e. having assay errors is effectively the same as having a smaller, less informative sample of seeds). Compensating for such errors may require prohibitively large seed samples.

Assay types and statistical implications

Several diagnostic methods have been developed to detect specific transgenic events (e.g. Monsanto's Yieldgard® event) or specific genes encoding transgenic traits (e.g. insect resistance) in seeds. In general, these assay methods fall into two categories. The first category includes methods that detect a specified protein or DNA target directly and attach a signal. Amplification of that signal enables detection of the protein target. The second class of assays requires amplification of a specific target first, with subsequent direct or indirect detection of the final product. The robustness and reproducibility of the diagnostic method used in seed testing is an important consideration when designing an appropriate testing plan. If the diagnostic assay has not been demonstrated to be accurate, estimates on seed lot purity may not be achieved. By understanding how diagnostic tests work, one can minimize false-positive and false-negative test results that may occur when testing pools of seed.

The three most widely used diagnostic tests for determining the presence or absence of specific genetic traits in seeds are enzyme-linked immunosorbent assays (ELISA) (Mason, 1992), lateral flow immunoassays (Hermanson *et al.*, 1992) and polymerase chain reaction (PCR) (Innis *et al.*, 1990). The first two assays detect proteins that confer a specific transgenic trait (e.g. insect resistance) and are examples of signal amplification methods. PCR is a DNA-based method and can be designed to detect either a unique event (e.g. Yieldgard®), specific genes conferring transgenic traits (e.g. an insect resistance gene) or specific genetic elements present in a transgenic crop (e.g. 35S promoter). PCR is an example of a method that involves target amplification.

ELISA assays and lateral flow immunoassays each require the development and use of specific antibodies or 'sticky fingers' that enable detection of a protein target derived from the insertion of a gene encoding a particular transgenic trait. A portion of the antibodies used in either the ELISA or the lateral flow immunoassay are designed to bind to the unique protein target. Attached to these antibodies are signals that can be amplified through a variety of chemical techniques. This amplified signal can then be easily detected if the specific protein is present above a certain threshold or level of detection (LOD). In the case of the ELISA assay, the amplified signal is generated in a liquid (typically in a 96-well plate

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

format) and can be measured against a known protein standard included in the assay so as to determine the relevant amount of target protein present in the sample. For the lateral flow immunoassays similar 'sticky fingers' antibodies are used in the test but the entire assay, including the signal amplification, takes place on a solid support – a special strip of paper. Using the lateral flow immunoassays one can easily obtain qualitative results (or, in other words, determine whether a particular protein target is present or not) if the protein is present above a defined threshold or LOD. Semi-quantitative results can be achieved by comparing lateral flow immunoassays to known protein standards in controlled testing.

The sensitivity of both the ELISA assay and the lateral flow immunoassays is limited by the quantity of protein in a sample, the quality of the protein extracted from the seed, the specificity of the antibodies used and the type of signal amplification chemistry employed. False-positive results are generally due to antibodies that are too 'sticky'; therefore, they are not specific enough. False-negative results can be generated as a result of antibodies not being 'sticky' enough, when low amounts of the specific protein target are expressed in a sample, inhibitors to the binding of antibodies are present in the protein extract, and reagents fail during protein extraction or signal amplification steps (Mason, 1992). For both types of protein assays, the level of sensitivity would need to be determined based on the size of the pool of seeds as well as the type of seed.

PCR assays require the development and use of a pair of specific, small pieces of DNA (primers) or DNA 'parentheses'. These primers demarcate a DNA sequence that is the specific target to be amplified. Only the DNA target surrounded by the 'parentheses' should be specifically amplified. Once the DNA target is amplified and enough copies of that target are generated, it can be detected by a variety of techniques. The target can be extremely specific, for example to the Yieldgard® corn event, by choosing DNA primers that amplify the junction regions between the inserted DNA and the endogenous plant DNA. A more generic PCR assay may be developed by choosing to amplify a specific gene in corn, such as a gene that would confer insect resistance, that is shared by many corn events, such as Monsanto's and Novartis' insect resistance corn products. PCR can provide both qualitative and quantitative results, depending on the type of analysis performed. PCR with an end-point determination of the final amplified product can provide qualitative information, while monitoring the accumulation of PCR product during the amplification process, or 'real time' monitoring, yields quantitative results when compared to known DNA standards. PCR is an extremely sensitive technique, as the amount of target amplification that could occur in a PCR reaction is theoretically unlimited. Up to 1 billion-fold amplification or higher can be achieved. Methods, like PCR, that utilize target amplification can be coupled with detection methods that utilize signal amplification to increase sensitivity over non-signal amplification detection methods. Despite the high level of sensitivity attributed to PCR, the optimal seed pool size would need to be determined for every assay in order to ensure that the assay can perform reliably in the seed testing plan.

False-positive PCR results can be attributed to non-specific DNA 'parentheses' and therefore amplification of the wrong or multiple targets; or contamination of a sample with low levels of the DNA target from other sources (Innis *et al*., 1990). For example, conventional corn kernels could be exposed to grain dust or plant residue from transgenic plants. Even DNA targets previously amplified by PCR tests on prior samples can float on dust in the laboratory and potentially taint new samples, thereby generating false-positive results. False-negative results in PCR tests can be due to poor quality or degraded DNA, the presence of PCR inhibitors from DNA extracted from plant or seed material, incomplete target amplification, reagent failure during subsequent DNA extraction and/or target amplification steps, and from reaction conditions that are insufficiently stringent. False-negative results due to a low quantity of DNA target alone are not likely since PCR amplification can usually overcome this problem. However, excessive amounts of total, non-target DNA included in a PCR reaction can cause inhibition, resulting in poor amplification of low-abundance target DNA templates.

Accurate diagnostic tests for seeds are important in developing an appropriate seed testing plan. With knowledge of how to control diagnostic assay performance, the assay error rate can be minimized for ELISA, lateral flow immunoassays and PCR. When the assay error rate has been determined for a particular diagnostic test, this information should be included in the overall statistical calculation to ensure realistic estimates of producer's and consumer's risks are obtained for a given testing plan.

Testing plan selection

Several options exist for designing a seed lot testing plan that considers the testing objective as well as budget and time resources. Key inputs such as the LQL, AQL and known assay system error rates must be available to design a realistic testing plan with respect to consumer's and producer's risks. This section presents some options for the design of testing

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

plans. The use of seed pools and double-stage testing plans are presented as options to reduce the time and cost of testing in some situations. The impact of false-positive and false-negative rates on testing plan results is also discussed in this section.

Testing seed pools

Seed testing can be expensive. For example, testing 400 individual seeds for the presence or absence of a transgenic event or trait can be labour intensive, and it requires a considerable amount of reagents. Often, resources can be saved by testing pools of seeds rather than individual seeds. For example, 400 individual seeds can be divided into 40 seed pools of 10 seeds each. Each seed pool can be ground into a homogeneous mix of flour, which is tested for the transgenic trait rather than testing each individual seed. This is a tenfold reduction in the cost and number of assays needed. Figure 5 shows that, with a rejection criteria of greater than or equal to four seeds or seed pools, testing plan properties are very similar for both the individual seed testing plan and the seed pool testing plan described here (i.e. consumer's and producer's risks will essentially be the same).

Due to the qualitative nature of most high-throughput protein and DNA assay test systems, seed pooling is most efficient when the absence of a particular event or trait is the 'pure state' and deviants show up as the impure presence in the pool. If the testing objective is to ensure that a particular transgenic trait is present at some high purity level (e.g. transgenic seed lot purity testing), then individual seeds may need to be tested. When testing seed pools, many currently available assay systems provide only qualitative results (i.e. absence versus presence), so that the number of seeds in a pool with a particular transgenic trait cannot be estimated. The qualitative assay is only able to determine if one or more seeds in the pool have the transgenic trait.

To reduce costs, it may seem advantageous to use testing plans with very large seed pools, thus requiring fewer assays to be run. These types of testing plans should be used with caution as they have the potential of a very high producer's risk. In addition, these testing plans may have high assay system error rates, since a potential protein or DNA target will be diluted in large seed pools. These error rates can have a large impact on the consumer's and producer's risk (especially when only a few large pools are tested) and should be considered carefully along with potential testing plans prior to settling on a given pool size. Examples of how to incorporate assay system error rates into testing plans are provided later. Figure 6 shows an example of two testing plans that use seed pooling. One testing plan tests 60 seed pools, each comprised of 50 seeds, and tolerates up to 17 deviant pools; the other plan tests six seed pools, each comprised of 500 seeds, and tolerates up to five deviant pools. Both plans test a total of 3000 seeds. Both testing plans have an AQL at 0.5% impurity and a LQL at 1.0% impurity. The testing plan with pools of 50 seeds has a consumer's risk at about 5% and a producer's risk at about 10%.

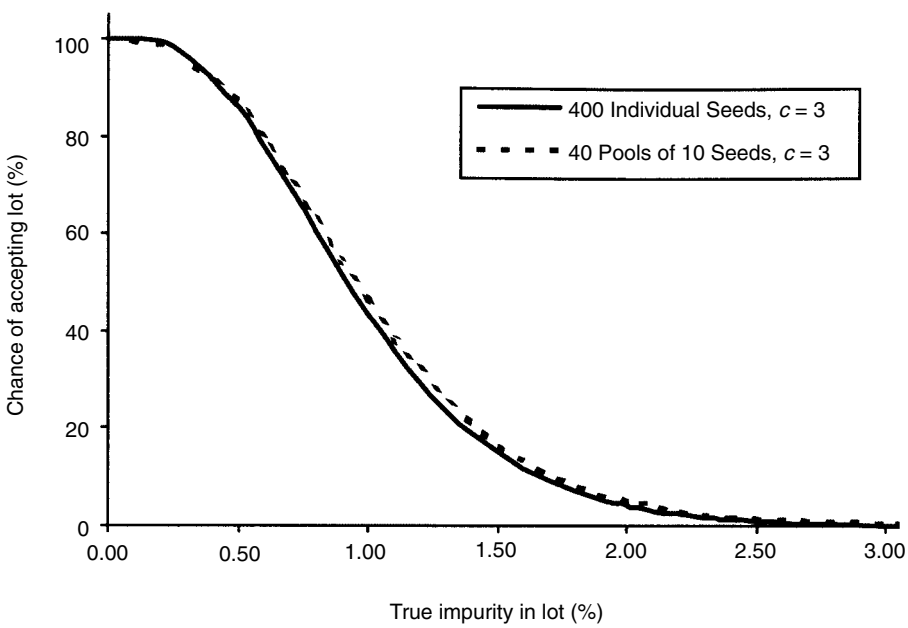


Figure 5. Virtually identical operating characteristic curves for an individual seed testing plan and a seed pool testing plan. Note that the total number of seeds tested remains the same.

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

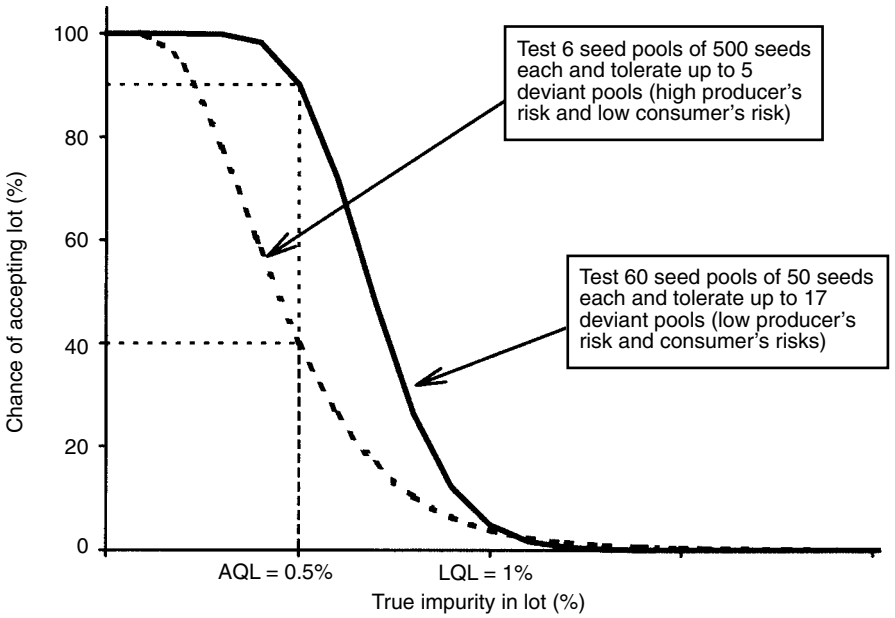


Figure 6. Operating characteristic curves for two large seed pool testing plans. The testing plan with the pool of 500 seeds has more difficulty reducing producer's risk levels.

The testing plan with pools of 500 seeds also has a consumer's risk of about 5%, but the producer's risk is high, at about 60%.

The testing plan with pools of 500 seeds in Fig. 6 may be a reasonable plan if the AQL is actually much lower (i.e. the AQL and LQL are further apart). In the previous example the AQL was one-half the LQL (LQL = 1% and AQL = 0.5%). If the AQL was one-tenth the LQL, then the large seed-pool plan may be reasonable in terms of the risks involved. If the AQL was 0.1%, the producer's risk would be reduced from 60% to less than 1%. To change the AQL to this level, seed producers would need to be confident that the majority of production seed lots to be tested are at 0.1% impurity (99.9% pure) or better.

It is also important to note that a testing plan with large seed pools may appear to have non-intuitive rejection criteria. For example, the testing plan in Fig. 6 that is based on six pools of 500 seeds has a low consumer's risk, even though it will result in the acceptance of seed lots when as many as five of the six seed pools are classified as deviant. While this rejection criterion may be initially unsettling to consumers, the reason why this testing plan will accept seed lots with so many deviant seed pools can be explained. If the impurity level in a lot is at the LQL (for example 1% impurity), it is expected that a pool of 500 seeds would have five deviant seeds on average and would thus be tested as deviant (one or more seeds have the impurity characteristic). This means that one would expect the majority of seed pools to test as deviant even when the true impurity in the lot is at, or slightly below, 1%, the LQL. The key information to make the decision that the seed lot impurity is below a 1% threshold is found in the single non-deviant seed pool rather than the five deviant seed pools. The probability of observing even a single non-deviant seed pool, given that the true lot impurity is 1%, is very small. This provides the strong evidence that the true lot impurity is below the 1% threshold.

Single-stage and double-stage testing plans

The number of stages that are conducted during seed lot testing can impact cost and other testing resources. A single-stage testing plan, as its name implies, has only one testing stage. This means that samples are taken from the lot, evaluated once, and a decision is made based on the results to accept or reject the seed lot. All of the testing plan examples that have been discussed in the previous sections have been single-stage testing plans.

In a single-stage testing plan, a specified number of individual seeds or seed pools are selected randomly from the sample, tested, and the number of deviant seeds or seed pools is observed. If the number of deviants detected exceeds the maximum number of deviants specified in the plan, then the lot is rejected. Otherwise, the seed lot is accepted.

A double-stage testing plan is somewhat more complicated than a single-stage testing plan, but may

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

result in a reduction of both cost and testing resources. A double-stage testing plan is generally set up so that fewer seeds are tested in the first stage than in the second stage. This is done to maximize the amount of sampling and testing resources that can potentially be saved. Initial test aliquots (i.e. subsamples of seeds) are taken from the sample and evaluated. Based on this evaluation, three different decisions can be made: (1) accept the lot; (2) reject the lot; or (3) take a second set of seed subsamples from the sample and retest. The test results from the first and second stages of testing are combined and used to determine whether the seed lot should be accepted or rejected.

Since three possible decisions can be made after the first stage of testing, and two after the second stage, there needs to be a different set of acceptance/rejection criteria than those used in single-stage testing plans. The first decision is based on the acceptance criterion of the first stage of sampling, which is given by c_1, the maximum number of deviants that result in acceptance of the seed lot after the first stage of testing without requiring the second stage of testing. The second decision is based on the rejection criterion, which is given by c_2, the number of deviants that, if exceeded in the first stage, will result in rejection of the seed lot. If the number of deviants after the first stage of sampling is greater than c_1 but does not exceed c_2, then a second stage set of samples is taken for evaluation. The third and final decision is then made based on the total number of deviants found in the first and second stages combined. If this total exceeds a threshold c_3 ($c_3 \geq c_2$), the seed lot is rejected. Otherwise the lot is accepted. For the two-stage plans described in many standard quality control texts (e.g. Montgomery, 1997), the first- and second-stage rejection thresholds are restricted to be equal (i.e. $c_3 = c_2$). For the sake of consistency with these texts, this simplification is also used in this paper. The use of a single rejection threshold simplifies the testing plan and reduces the effort involved in finding and tabulating optimal two-stage plans. It should be noted, however, that improvements in efficiency can sometimes be achieved when c_3 is allowed to exceed c_2.

A double-stage testing plan is very cost effective if the majority of the seed lots to be tested are either at a very high or a very low level of genetic impurity relative to the AQL and LQL. The seed lots that fall in either of these two categories can generally be classified (accepted or rejected) in the first stage of testing, thus saving sampling resources and time. The only seed lots that would require the second stage of testing are those that are close to the threshold (i.e. seed lots that are close to or within the range between the AQL and LQL). On the other hand, if the majority of seed lots are expected to be close to the LQL or AQL, then a double-stage testing plan might require the retesting of most lots, which may not be cost effective. The expected cost of testing, EC, for a plan that requires a sample size of n_1 at the first stage and n_2 at the second stage might be modelled simply as:

$$\mathrm{EC} = Q_1 + n_1 B_1 + P_{\mathrm{retest}} (Q_2 + n_2 B_2 + T)$$

Here Q_1 and Q_2 are the fixed costs of conducting a seed test at the first and second stages, respectively. B_1 and B_2 are the testing costs per seed (or seed pool) for the two stages. P_{retest} is the probability of a retest, and T is the cost incurred by having to wait for a second set test to make a final decision about the seed lot. Often, but not always, $Q_1 = Q_2$ and $B_1 = B_2$. If the greatest costs are those per seed or seed pool tested (i.e. B_1 and B_2), then the expected cost of a second-stage design will not usually exceed that of a single-stage design – even when P_{retest} is near 1. However, if each new test requires costly preparations (Q_1 and Q_2), or if delays have financial consequences (T), then two-stage testing may be less efficient when the probability of retest is high.

Figure 7 shows an OC curve for a double-stage testing plan. The dashed curve in the figure gives the probability that a lot will need to be retested in a second stage of this testing plan, given the true lot impurity. This curve shows that lots that have true impurities less than 0.5% or greater than 3%, for example, can usually be classified in the first stage of the plan. The only seed lots that are likely to require further testing in a second stage would be those seed lots that are close to, or within, the range of 0.5–3% impurity. This testing plan would be very efficient if the seed lots to be tested are generally outside of this range.

Table 2 contains examples of five testing plans. The LQL and AQL for each testing plan are set at 1% and 0.5% impurity, respectively, for this example. The first, second and fifth testing plans are single-stage plans, and the third and fourth plans make use of double-stage testing.

The first testing plan involves testing 3000 seeds individually (i.e. pool size is equal to 1) and accepts seed lots if 21 or fewer seeds test as deviants. Although this type of a testing plan is economically unrealistic for most assays, it is used here as a baseline for comparison with the other four testing plans. The producer's and consumer's risks for this testing plan are both at 5% (i.e. 95% producer's and consumer's confidence in the decision).

The second testing plan in Table 2 is a single-stage plan with 60 seed pools of 50 seeds each. Here again, 3000 seeds are evaluated, but the number of assays is reduced from 3000 to 60. The consumer's risk is still at 5% but the producer's risk is 10%. Ten per cent of the time, seed lots with 0.5% impurity will be rejected using testing plan 2.

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

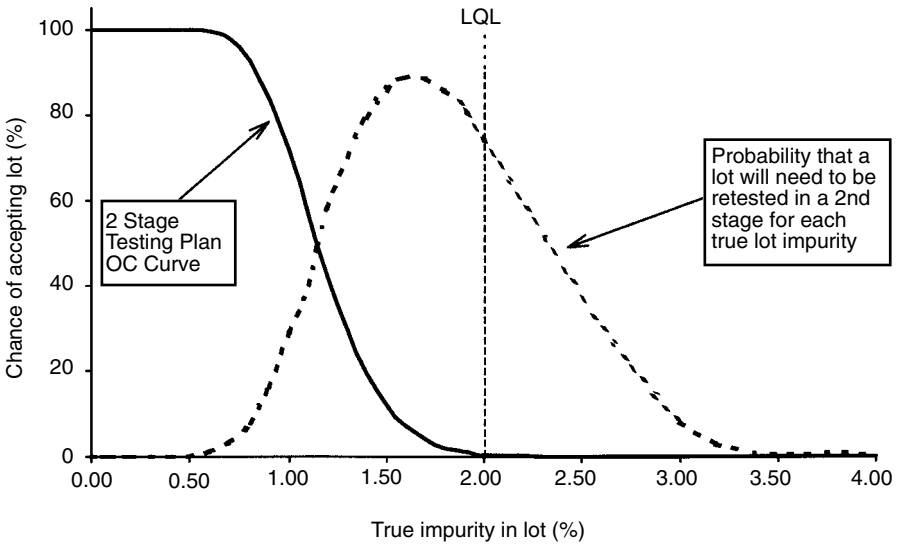


Figure 7. Operating characteristic and retesting probability curve for a two-stage testing plan. Note that lots with true impurity below 0.5% or above 3% will rarely need to be retested in a second stage.

Table 2. Seed testing plan examples (LQL = 1%, AQL = 0.5%)

Testing plan	Number of assays: First stage	Number of assays: Second stage	Seed pool size	c_1*	c_2*	Producer's risk (%)	Consumer's risk (%)	Retest rate at AQL (0.5%) (%)	Retest rate at 0.1% impurity (%)
1	3000	–	1	–	21	5	5	–	–
2	60	–	50	–	17	10	5	–	–
3	30	30	50	5	17	10	5	68	0.3
4	30	60	50	6	26	5	4	51	0
5	10	–	300	–	7	61**	1	–	–

*c_1 and c_2 are the acceptance/rejection criteria for the testing plan. If the number of deviants (seeds or seed pools) does not exceed c_1 after the first stage of sampling, then the lot is accepted. If the total number of deviants exceeds c_2 after the first or second stages, then the lot is rejected.
**This producer's risk is reduced to 3% if the AQL = 0.2%, which would make this a reasonable testing plan.

The third testing plan in Table 2 makes use of double-stage testing. Thirty seed pools of 50 seeds each are evaluated in the first stage of sampling. If five or fewer of these seed pools are deviants, then the seed lot is accepted and no more testing is needed. If more than 17 deviant pools are observed, then the lot is rejected after the first stage. If the number of detected deviant seed pools is from 6 to 17 in the first stage, then an additional 30 seed pools of 50 seeds each are evaluated. If the number of deviant seed pools from the first and second stages combined exceeds 17, then the lot is rejected; otherwise the lot is accepted after the second stage of testing. The producer's and consumer's risks for testing plan 3 are equivalent to those of the second testing plan (10% and 5%, respectively). However, with the third plan, 68% of the lots with impurity levels at 0.5% would need to be retested. This testing plan is very economical if most of the seed lots are in the range of less than 0.2% or greater than 3% true impurity (i.e. the retest rate is less than 6%). If this is the case, then most of the seed lots will be classified in the first stage, and half of the testing resources are saved for these lots. If most of the seed lots have impurity levels in the range 0.5–3.0%, then this testing plan would not provide a substantial savings of testing resources over the single-stage sampling plan.

The fourth testing plan in Table 2 tests 30 seed pools of 50 seeds each in the first stage and 60 seed pools of 50 seeds each in the second stage. A lot is accepted in the first stage if no more than six deviant pools are detected; a lot is rejected if more than 26

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

deviants are detected in the first and second stages combined. This testing plan will accept about 50% of the seed lots with true impurity at 0.5% or less in the first stage. This testing plan has producer's and consumer's risks that are much closer to the baseline individual seed testing plan described in plan 1 (i.e. producer's risk is 5% and consumer's risk is 5%).

The fifth testing plan in Table 2 tests 10 pools of 300 seeds each in a single-stage plan and accepts a lot if seven or fewer deviants are found. This testing plan has a very reasonable consumer's risk of 1%; however, the producer's risk is 61%. This is due to the very large size of the seed pools tested compared to plans 2, 3 and 4. However, this testing plan could be reasonable if the AQL could be changed from 0.5% to 0.2%, in which case the producer's risk is reduced from 61% to 3%.

There are other types of seed testing plans in addition to single- and double-stage testing plans. An example is a sequential testing plan. In sequential testing plans, the number of seeds or seed pools to be tested is not specified. An individual seed or seed pool is tested. After each sample is tested, the information from all the samples assayed so far is evaluated to determine if the lot can be classified. If the lot cannot be classified, then sample testing continues. The samples continue to be tested until enough information is available to classify the lot. In such schemes, the decision rule is established with respect to the AQL and LQL values prior to testing. Sequential testing plans are not discussed in more detail since the authors expect that they are not practical for most seed lot testing situations.

Assay system error impact

All of the testing plans that have been discussed in examples so far have assumed that the assay system false-positive and false-negative error rates are zero. Since this is generally not the case, this section discusses briefly the impact of these error rates on testing plan results. Assay system error rates need to be determined for a given seed pool size before a reliable testing plan using that pool size can be developed.

As one would expect intuitively, the false-negative rate adversely affects consumer's risk and the false-positive rate adversely affects the producer's risk. The magnitude of these effects depends on the number of assays and the magnitude of the error rates themselves. Table 2 shows that testing plan 2 has a consumer's risk of 5%, assuming the error rates are zero. If the false-negative rate for the assay system were really 5%, then the consumer's risk would increase to 9%. If the false-negative rate were really 10%, then the consumer's risk would be inflated to 15%. The Appendix contains formulas that incorporate assay errors into testing plans.

In some cases testing plans can be developed that adjust for 'low' error rate levels and still maintain reasonable consumer's and producer's risks. However, this assumes that these error rates are known. Laboratory testing should be conducted to provide estimates of these error rates before testing seed lots. A ring (multi-laboratory) test may be conducted which involves several testing facilities to obtain some reasonable estimates of false-positive and false-negative rates. Ring tests require substantial resources and effort in order to obtain reasonable estimates of assay error rates.

Lot impurity estimation

The primary objective of seed lot acceptance sampling is to classify lots into 'accept' or 'reject' categories. In this context, estimating the level of impurity in the seed lot is of secondary importance. In seed lot acceptance sampling, the focus is on determining whether the level of impurity in a given seed lot is above or below a given threshold. Seed lot impurity estimation seeks to estimate a realistic range or confidence interval in which the true impurity of the seed lot is found. Note that for seed lot impurity estimation, there is no requirement for a predetermined threshold.

There is a close relationship between seed lot classification and seed lot impurity estimation, which will be discussed using an example. Suppose that a seed producer wishes to guarantee that seed lots sold have less than 2% trait presence with 95% confidence. In seed lot classification, one would simply make the decision to reject or accept the seed lot based on the testing results. If the seed lot is accepted, a statement is made that there is 95% confidence that the true presence of transgenic traits in the seed lot is below 2%. This statement also implies that the upper confidence limit of the estimated impurity is equal to or less than 2%. If, on the other hand, the seed lot impurity estimation was the focus, then seed lot impurity would first be estimated based on the formulas discussed in the Appendix. This estimate is the expected impurity in the seed lot based on testing results. A 95% upper confidence limit for this impurity estimate could then be calculated. The true impurity in the seed lot would be expected to be no worse than this limit with 95% confidence. If the upper confidence limit is above the 2% threshold, then the seed lot is rejected. If the upper confidence interval is below the 2% threshold, then the seed lot is accepted.

The seed purity estimation approach provides more information, since the 95% upper confidence limit is calculated rather than stating whether the upper 95% confidence limit is above or below the 2%

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

threshold. There may be two seed lots that are both accepted using these accept/reject criteria, so the decision is made that their impurity is below 2%. However, calculating the 95% upper confidence limit may show that one seed lot upper limit was 1.9%, while the other seed lot upper limit was 0.5%. Hence seed lot impurity estimation, although requiring more calculations, yields more information than seed lot classification. Methods for calculating confidence limits for the level of impurity are given in the Appendix.

Final comments and implementation

We hope that this paper has given the reader a better understanding of some statistical aspects of seed lot acceptance sampling. The concepts and direction given in this paper can be used to help design appropriate testing plans that minimize the risk to consumers and to seed producers.

A number of different software packages exist that can aid in setting up testing plans. The calculations can be programmed in a spreadsheet application using some existing functions augmented by a few user-defined functions. The authors programmed the formulas from the Appendix into a Microsoft Excel® spreadsheet application for use with the testing plan calculations in this paper, as well as, in practice, to assist clients with designing appropriate testing plans. This spreadsheet can be downloaded from the SeedQuest® website (http://www.seedquest.com/best/spreadsheet). Figure 8 shows Testing plan 2 from Table 2 (modified to include an assay false-negative rate of 1%), to illustrate how this spreadsheet is used. The following testing plan information is entered: testing plan type (double or single stage), number of tested pools ('# of Batches'), pool size, LQL, AQL and accept/reject criteria. The spreadsheet then calculates the OC curve as well as the producer's and consumer's risk at the AQL and LQL, respectively. This spreadsheet also allows the user to enter 'known' assay error rates to see their impact on producer's and consumer's risk for a given testing plan. A 1% false-negative rate is entered into the spreadsheet for this example to show the impact on the producer's and consumer's risks. The risk differences in Table 2 and Fig. 8 for Testing plan 2 can be attributed to the 1% false-negative rate that was input.

The Grain Inspection, Packers and Stockyard Administration (GIPSA) division of USDA has provided a spreadsheet application on their web site that has similar functionality to the spreadsheet that the authors have written. The web site address is *http://www.usda.gov/gipsa/biotech/biotech.htm*. This application is called *Sample Planner* and is a Microsoft Excel® spreadsheet. Although this application was written for the grain testing industry, it will still be quite useful for the commercial seed testing industry and seed science researchers. In its current form, this spreadsheet can be used to calculate producer's confidence (i.e. 1 – producer's risk) in single-stage testing plans. Also by entering an LQL in place of the AQL, the 'probability of accepting the AQL' becomes the consumer's risk.

The validity of the sampling and testing plan assumptions discussed in this paper are the keys to the validity of the decisions that are made to accept or

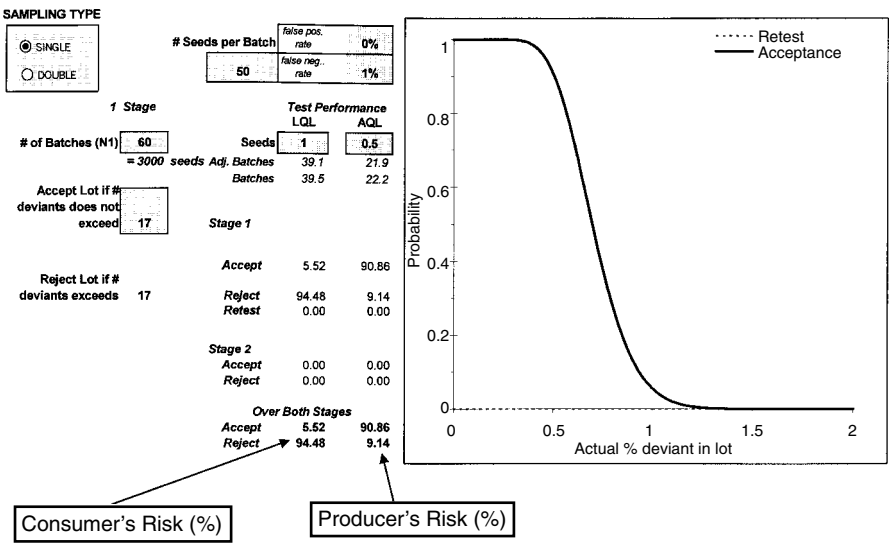


Figure 8. Example of a spreadsheet implementation of formulas in the Appendix. Testing plan 2 in Table 2 is used in this example. A 1% false-negative rate is input for this example.

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

reject seed lots based on purity testing. If these assumptions are violated, the consumer's risk and/or producer's risks associated with the decisions may be misleading. Because of their importance, these key assumptions are reiterated here.

- *The purity of the lot is relatively homogeneous across the lot*. Seed lots should be organized so that there is less variability in quality within lots than there is between lots. For example, a single lot should contain seeds from producers in the same geographical area rather than mixing seeds from different areas. Not only does this make the definition of a lot coherent, it makes testing more robust compared to non-simple random sampling.
- *The sample quality characteristics are representative of the lot quality characteristics.* A sampling plan should be used that mimics, as closely as possible, a simple random sampling plan. Also, when seed pools are used in testing plans, the ground flour must be a homogeneous mix so that the aliquot taken for the assay is typical of the characteristics of the whole seed pool.
- *Assay system false-negative and false-positive rates are relatively low.* The validity of this assumption needs to be considered carefully. For example, testing plans that have particularly large seed pools may have unacceptably high false-negative rates. Very sensitive testing methods, such as those based on polymerase chain reaction (PCR), which has the potential to amplify DNA targets up to 1 billion-fold, are susceptible to generating false-positive results. The presence of trace levels of deviant seeds from other sources, either in the samples or in a testing laboratory, could also carry the specific target intended to be identified in certain PCR assays and cause a false-positive test result (e.g. viral and bacterial contamination in seeds cross-reacting with 35S promoter PCR tests). Error rates should be quantified empirically, with adjustments made to the testing plan to compensate for these error rates and still maintain acceptable producer's and consumer's risks.

Appendix

This section contains the formulas that are used to calculate producer's and consumer's risks for the seed testing plans discussed in this paper. Formulas for estimating lot impurity based on the sampling and testing results are also given for single-stage sampling plans for the case of testing individual seeds as well as seed pools.

If a seed lot is large relative to the number of seeds that are tested from the lot (i.e. tested seed is less than 10% of the lot), then the number of deviant individual seeds or deviant seed pools in a sample taken from this lot has a binomial distribution with parameters n and p_b, where n is the number of individual seeds or seed pools tested and p_b is the probability that an individual seed or seed pool is deviant. The probability calculations used in this application are based on standard lot acceptance sampling techniques using the binomial distribution. These basic lot acceptance sampling methods have been used in industry for many years. Montgomery (1997) is one of many textbooks that contains a discussion of acceptance sampling plans. The acceptance sampling approach outlined in this reference is extended in this paper to include: (1) testing seed pools rather than individual seeds only, and (2) incorporating assay system error rates (i.e. false-positive and false-negative rates) into the calculation of consumer's and producer's risks for seed testing plans.

A few parameters need to be defined before the risk formulas can be derived: p is the true unknown impurity in the seed lot; n, the number of individual seeds or seed pools to be tested; m, the number of individual seeds in a seed pool (if seeds are tested individually, then $m = 1$); δ, the false-positive rate of the assay method; and λ, the false-negative rate of the assay method.

The probability that there is at least one deviant seed in a given seed pool or batch is equal to 1 minus the probability that there are no deviant seeds in the pool (P). This probability is calculated as a function of p and m as

$$p_b = 1 - P = 1 - (1 - p)^m.$$

The parameter p_b gives the probability that a given seed pool is a true deviant (i.e. true positive). Note that if m is equal to 1 (the case of individual seed testing), then p_b is equal to p. When independent seed pools are used in a testing plan, the number of deviant seed pools follows a binomial distribution with parameters n and p_b. This formulation takes into consideration the limitation of the qualitative nature of the assaying systems presently used to test seeds, such as event-specific PCR (i.e. the method can only detect the presence of one or more deviant seeds in the pool). This limitation will likely change in the future.

The remainder of the discussion in this appendix will present formulas for testing seed pools. However, if individual seeds are tested, these formulas can be used by taking $m = 1$.

The impact of assay method false-positive and false-negative rates on the consumer's and producer's risks must be considered. The parameter p_a is defined to be equal to the probability that a given seed pool tests deviant (i.e. tests positive) using the assay method. Note that even though a seed pool tests deviant, it may not necessarily be a true deviant when the values of δ and λ are non-zero. The parameter p_a

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

can be expressed as a function of the false-negative and false-positive rates as:

$$\begin{aligned} p_a &= P(\text{test deviant}) = P(\text{true deviant}) \\ &\quad + P(\text{false deviant}) \\ &= P(\{\text{test deviant}\} \cap \{\text{deviant}\}) \\ &\quad + P(\{\text{test deviant}\} \cap \{\text{not deviant}\}) \\ &= P(\text{test deviant} \mid \text{deviant})P(\text{deviant}) + P(\text{test} \\ &\quad \text{deviant}) \mid \text{not deviant})P(\text{not deviant}) \\ &= (1 - \lambda)p_b + \delta(1 - p_b) \end{aligned}$$

Single-stage testing plans

Two quantities that still need to be defined in order to calculate the consumer's and producer's risks for a single stage-testing plan are c, the maximum number of seed pools that can test deviant and result in acceptance of the seed lot, and d, the observed number of seed pools that test deviant from a sampled seed lot.

The reader may refer to the section of this paper on seed lot testing uncertainty to review the definitions for LQL, AQL, consumer's risk and producer's risk, which are used in the formulas that follow. Here the LQL and AQL are written in terms of impurity level (1 – purity level). The probability that a lot will be accepted given its true lot impurity level (p) is at the LQL, δ and λ is given by

$$P\left(\text{Accept Lot} \middle| p = \text{LQL},\ \delta,\ \lambda\right) = P\left(d \le c \middle| p = \text{LQL},\ \delta,\ \lambda\right)$$
$$= \sum_{i=0}^{c} \binom{n}{i} p_a^i \left(1 - p_a\right)^{n-i}.$$

This is the formula used to calculate the consumer's risk for single-stage sampling plans. In this formula and the formulas that follow, it is helpful to remember that p_a, the probability that a seed pool tests deviant, is a function of p (the underlying lot impurity), m (the number of seeds in a pool), and the false-positive and negative rates of the assay method. Note that testing plans using seed pooling ($m > 1$) are useful for testing for unintended events only, since all assay systems presently used are qualitative.

The probability that a lot is rejected given that its true lot impurity level (p) is at the AQL, δ and λ is given by

$$P\left(\text{Reject Lot} \middle| p = \text{AQL},\ \delta,\ \lambda\right) = P\left(d > c \middle| p = \text{AQL},\ \delta,\ \lambda\right)$$
$$= \sum_{i=c+1}^{c} \binom{n}{i} p_a^i \left(1 - p_a\right)^{n-i}.$$

This is the formula used to calculate the producer's risk for single stage sampling plans.

Double-stage testing plans

The calculation of producer's and consumer's risks becomes somewhat more complicated for double-stage testing plans than for single-stage testing plans. There are a few additional terms that must be defined before probability calculations can be made for this type of testing plan. They are: n_1, the number of independent seed pools to be tested in the first stage of a double-stage testing plan; n_2, the number of independent seed pool samples to be tested in the second stage of a double-stage testing plan; c_1, the maximum number of allowable deviant seed pools for acceptance of the seed lot in the first stage of a two-stage testing plan; c_2, the number of deviant seed pools that, if exceeded in the first stage, will result in rejection of the seed lot; c_3, the maximum number of allowable deviant seed pools in the first and second stages combined for acceptance of the seed lot in the second stage; d_1, the number of deviant seed pools in the first stage of sampling; and d_2, the number of deviant seed pools in the second stage of sampling.

In most seed testing situations and in this paper, we let n_1 be strictly greater than c_2. There are some double-stage testing situations where c_2 is greater than n_1, such as the case when seed lots are not allowed to be rejected in the first stage. To calculate the producer's and consumer's risks associated with a double-stage testing plan, there are a few intermediate probability calculations that must first be made. The probability of accepting a lot with impurity level p, false-positive rate δ, and false-negative rate λ in the first stage of sampling is equal to

$$P\left(d_1 \le c_1\right) = \sum_{i=0}^{c_1} \binom{n_1}{i} p_a^i \left(1 - p_a\right)^{n_1 - i}.$$

The probability of rejecting a lot with impurity level p, false-positive rate δ, and false-negative rate λ in the first stage of testing is equal to

$$P\left(d_1 > c_2\right) = \sum_{i=c_2+1}^{n_1} \binom{n_1}{i} p_a^i \left(1 - p_a\right)^{n_1 - i}.$$

The probability that there needs to be a retest of a lot with impurity level p (i.e. go to the second stage of testing) is equal to

$$P\left(c_1 < d_1 \le c_2\right) = \sum_{i=c_1+1}^{c_2} \binom{n_1}{i} p_a^i \left(1 - p_a\right)^{n_1 - i}.$$

The probability of accepting a seed lot in the second stage of sampling is equal to the probability of retesting and accepting the seed lot in the second stage of sampling. Since these two events are not independent, the simple product of these two

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

probabilities cannot be taken. The correct probability of accepting a seed lot in the second stage of sampling is equal to

$$P\left(\left\{c_1 < d_1 \le c_2\right\} \text{ AND } \left\{d_1 + d_2 \le c_3\right\}\right)$$
$$= \sum_{i=c_1+1}^{c_2} P\left(d_1 = i\right) P\left(d_1 + d_2 \le c_3 \middle| d_1 = i\right)$$
$$= \sum_{i=c_1+1}^{c_2} \left[\left\{ \binom{n_1}{i} p_{\mathrm{a}}^{i} \left(1 - p_{\mathrm{a}}\right)^{n_1 - i} \right\} \times \left\{ \sum_{j=0}^{c_3 - i} \binom{n_2}{j} p_{\mathrm{a}}^{j} \left(1 - p_{\mathrm{a}}\right)^{n_2 - j} \right\} \right].$$

Each element of the grand sum above is the probability of seeing no more than $c_3 - i$ deviants in the second sample, given that exactly i deviant seed pools are observed in the first stage of sampling, multiplied by the probability of seeing exactly i deviant seed pools in the first stage. Since we are focusing attention only on two-stage designs that have a single rejection threshold, c_2, the quantity c_3 will be set equal to c_2 in the sequel.

The probability of rejecting a seed lot in the second stage of testing is equal to the probability of retesting the lot and rejecting it in the second stage. This probability can also be expressed as

$$P\left(\left\{c_1 < d_1 \le c_2\right\} \text{ AND } \left\{d_1 + d_2 > c_2\right\}\right)$$
$$= \sum_{i=c_1+1}^{c_2} \left[\left\{ \binom{n_1}{i} p_{\mathrm{a}}^{i} \left(1 - p_{\mathrm{a}}\right)^{n_1 - i} \right\} \times \left\{ 1 - \sum_{j=0}^{c_2 - i} \binom{n_2}{j} p_{\mathrm{a}}^{j} \left(1 - p_{\mathrm{a}}\right)^{n_2 - j} \right\} \right].$$

Using the above probabilities for two-stage testing plans, the producer's and consumer's risk probabilities can be calculated. The producer's risk, or the probability that the seed lot is rejected given that the true impurity (p) is at the AQL (and given δ and λ), in a two-stage testing plan is equal to

$$P\left(\text{Reject Lot} \middle| p = \text{AQL},\ \delta,\ \lambda\right)$$
$$= P\left(\left\{d_1 > c_2 \text{ OR } d_1 + d_2 > c_2\right\} \middle| p = \text{AQL},\ \delta,\ \lambda\right)$$
$$= P\left(d_1 > c_2\right) + P\left(\begin{array}{l} \left\{c_1 < d_1 \le c_2\right\} \\ \text{AND } \left\{d_1 + d_2 > c_2\right\} \end{array} \middle| p = \text{AQL},\ \delta,\ \lambda \right).$$

The consumer's risk, or the probability of accepting the seed lot given that the true impurity (p) is at the LQL, in a two-stage testing plan is equal to

$$P\left(\text{Accept Lot} \middle| p = \text{LQL},\ \delta,\ \lambda\right)$$
$$= P\left(\left\{d_1 \le c_1 \text{ OR } d_1 + d_2 \le c_2\right\} \middle| p = \text{LQL},\ \delta,\ \lambda\right)$$
$$= P\left(d_1 \le c_1\right) + P\left(\begin{array}{l} \left\{c_1 < d_1 \le c_2\right\} \\ \text{AND } \left\{d_1 + d_2 \le c_2\right\} \end{array} \middle| p = \text{LQL},\ \delta,\ \lambda \right).$$

Lot impurity estimation

Although the primary objective of seed lot acceptance sampling is to classify lots, estimating the lot impurity and putting confidence limits on this estimate are often desired. Formulas for the lot impurity estimates will first be derived, followed by the associated formulas for the confidence intervals.

If individual seeds are tested (m = 1), then the estimate of lot impurity is simply calculated as

$$\hat{p} = \frac{d}{n}.$$

An approximate (1 − α) × 100 per cent upper confidence limit for lot impurity can be calculated using the equation

$$\hat{p}_{\mathrm{UL}} = \frac{\left(d+1\right) F_{1-\alpha,\,2d+2,\,2n-2d}}{\left(\mathrm{n} - \mathrm{d}\right) + \left(d+1\right) F_{1-\alpha,\,2d+2,\,2n-2d}}$$

when individual seeds are tested. The quantity F is the 1 − α quantile from an F-distribution with $2d + 2$ and $2n - 2d$ degrees of freedom. This equation comes from Johnson *et al.* (1993).

If $m > 1$, then $\hat{p}$ is an estimate of the proportion of deviant seed pools in the lot, rather than an estimate of the lot impurity. Recall from the earlier discussion that the probability a given seed pool is deviate is equal to

$$p_{\mathrm{b}} = 1 - \left(1 - p\right)^{m}$$

An estimate of p_{b} is given by

$$\hat{p}_{\mathrm{b}} = \frac{d}{n}.$$

The two previous equations can be set equal to each other to yield

$$\frac{d}{n} = 1 - \left(1 - p\right)^{m}.$$

This equation can be solved for p to give the following estimate for lot impurity when seed pooling is used, giving

$$\hat{p} = 1 - \left(1 - \frac{d}{n}\right)^{\frac{1}{m}}.$$

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

An approximate (1 – α) × 100 per cent upper confidence interval for lot impurity can be calculated by modifying the previous confidence limit to take into consideration seed pooling as

$$\hat{p}_{\mathrm{UL}} = 1 - \left(1 - \frac{(d+1)F_{1-\alpha,2d+2,2n-2d}}{(\mathrm{n}-\mathrm{d}) + (d+1)F_{1-\alpha,2d+2,2n-2d}} \right)^{\frac{1}{m}}$$

with the same F-distribution as stated for the individual seed testing case.

For two-stage plans, conventional calculations for an upper bound for the level of impurity are considerably more complex than is appropriate here. Fortunately there exists an extremely simple alternative based on a Bayesian approach with a uniform prior distribution (Lindley, 1970). This approach gives essentially the same results as the more complex conventional bound. This general (1 – α) × 100 per cent upper Bayesian confidence interval for lot impurity is given by

$$\hat{p}_{\mathrm{UL}} = 1 - \left[1 - Be_{1-\alpha,d,n-d} \right]^{\frac{1}{m}}$$

where the quantity Be is the 1 − α quantile from a beta distribution with parameters d and $n - d$. In this case, however, n and d are the total number of seeds tested and number of deviants, respectively, from all stages tested. Functions for the quantiles for both the F and the beta distributions are provided in most statistical and spreadsheet programs.

Acknowledgements

The authors thank the editor and a number of reviewers for input and suggestions that have enhanced the quality of this paper.

References

AOSA (1998) *Rules for testing seeds*. Columbus, Ohio, Association of Official Seed Analysts.

Cochran, W.G. (1977) *Sampling techniques* (3rd edition). New York, John Wiley.

Hermanson, G.T., Mallia, A.K. and Smith, P.K. (1992) *Immobilized affinity ligand techniques*. San Diego, Harcourt Brace Jovanovich.

Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (1990) *PCR protocols: A guide to methods and applications*. San Diego, Harcourt Brace Jovanovich.

ISTA (1999) *International rules for seed testing rules 1999*. Zurich, Switzerland, International Seed Testing Association.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1993) *Univariate discrete distributions* (2nd edition). New York, John Wiley.

Lindley, D.V. (1970) *Introduction to probability and statistics from a Bayesian viewpoint, part 2. Inference* (1st edition). Cambridge, Cambridge University Press.

Mason, M. M. (1992) *Immunochemical protocols. Methods in molecular biology*, Vol 10. Totowa, NJ, Humana Press.

Montgomery, D.C. (1997) *Introduction to statistical quality control* (3rd edition). New York, John Wiley.

USDA Grain Inspection, Packers and Stockyards Administration (1995a) *Mechanical sampling systems handbook*. Washington DC, USDA. (*http://www.usda.gov/gipsa*)

USDA Grain Inspection, Packers and Stockyards Administration (1995b) *Grain inspection handbook: Book 1*. Washington DC, USDA. *(http://www.usda.gov/gipsa*)

Received 31 August 2000
accepted after revision 14 December 2000
© CAB *International* 2001

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press

Grassland Ecophysiology and Grazing Ecology

Edited by **G Lemaire**, *INRA, France*, **J Hodgson**, *Massey University, New Zealand*, **A Moraes**, *Universidade Federal do Parana*, **P C de F Cavalho** *and* **C Nabinger**, *Universidade Federal do Rio Grande do Sul, Brazil*

October 2000 **432 pages** **Hardback**
ISBN 0 85199 452 0

£65.00 (US$120.00)

Readership: Plant ecology/physiology, grassland science, animal science.

The sustainability and ecophysiology of grazing systems are of growing importance. This research level book presents edited, key papers from the International Symposium on Grassland Ecophysiology and Grazing Ecology held in Curitiba, Brazil in August 1999. It considers how plants within grasslands respond to and are adapted to grazing animals.

This title:

- Is of worldwide interest, covering both temperate and tropical pastures
- Includes contributors and leading authorities from North and South America, Europe and Australasia

Contents:

- Sustainability of grazing systems: goals, concepts and methods, *J Hodgson and S C da Silva*

Part I: Environmental constraints and plant responses to defoliation

- Effects of nitrogen and water supply on N and C fluxes and partitioning in defoliated swards, *F Gastal and J L Durand*
- An integrated view of C and N uses in leaf growth zones of defoliated grasses, *H Schnyder, R Schäufele, R de Visser and C J Nelson*
- Effects of grazing on the roots and rhizosphere of grasses, *L A Dawson, S J Grayston and E Paterson*
- Reserve formation and recycling of carbon and nitrogen during regrowth of defoliated plants, *B Thornton, P Millard and U Bausenwein*

Part II: Morphogenesis of pasture species and adaptation to defoliation

- Shoot morphological plasticity of grasses: leaf growth *vs.* tillering, *C J Nelson*
- Tiller dynamics of grazed swards, *C Matthew, S G Assuero, C K Black and N R Sackville Hamilton*
- Effect of nitrogen on some morphogenetical traits of temperate and tropical perennial forage grasses, *P Cruz and M Boval*
- Modelling the dynamics of temperate grasses and legumes in cut mixtures, *J F Soussana and A Oliveira Machado*

Part III: Plant-Animal interactions

- Plant-animal interactions in complex plant communities: from mechanism to modelling, *I J Gordon*
- Modeling spatial aspects of plant-animal interactions, *E A Laca*
- Defoliation patterns and herbage intake on pastures, *M H Wade and P C de Carvalho*
- Selective grazing on grass-legume mixtures in tropical pastures, *C E Lascano*

Part IV: Sustainable grazing management of natural pastures

- Leaf tissue turnover and efficiency of herbage utilization, *G Lemaire and M Agnusdei*
- Dynamics of heterogeneity in a grazed sward, *A J Parsons, P Carrère and S Schwinning*
- Soil-plant-animal interactions and impact on nitrogen and phosphorus cycling and re-cycling in grazed pasture, *S C Jarvis*
- Sustainable management of pasture and rangelands, *J Stuth and G E Maraschin*

Part V: Problems of Animal Production Related to Pastures in Subtropical and Temperate Regions of South America

- *Campos* in Southern Brazil, *C Nabinger, A. de Moraes, and G E Maraschin*
- *Campos* in Uruguay, *E J Berretta, D F Risso, F Montossi, and G Pigurina*
- Argentina's Humid Pampa, *V A Deregibus*

For further information or to order please contact CABI *Publishing*, UK or an exclusive CABI *Publishing* distributor in your area.
Please add £2.50 per book postage and packing (excluding UK).

CABI Publishing, CAB International, Wallingford, Oxon, OX10 8DE,UK
Tel:+44(0)1491 832111 Fax:+44(0)1491 829292 Email:orders@cabi.org
CABI Publishing, CAB International, 10 East 40th Street, Suite 3203, New York, NY 10016, USA
Tel:+1 212 481 7018 Fax:+1 212 686 7993 Email:cabi-nao@cabi.org

https://doi.org/10.1079/SSR200166 Published online by Cambridge University Press