Zeitschrift: Mitteilungen aus Lebensmitteluntersuchungen und Hygiene = Travaux de chimie alimentaire et d'hygiène

Herausgeber: Bundesamt für Gesundheit

Band: 92 (2001)

Heft: 3

Artikel: Sampling procedures to determine the proportion of genetically modified organisms in raw materials. Part I, Correct sampling, good sampling practice

Autor: Lischer, Peter,

DOI: https://doi.org/10.5169/seals-981915

Sampling Procedures to Determine the Proportion of Genetically Modified Organisms in Raw Materials

Part I: Correct Sampling, Good Sampling Practice

Peter Lischer, Constat Consulting, Spiegel bei Bern

Received 4 January 2001, accepted 30 April 2001

Introduction

The use of genetically modified organisms (GMOs) in food products is becoming more widespread, and many consumers are concerned about their use in everyday products. The Swiss government revised its food regulation, introducing a threshold value of 1 % GMO content as the basis of food labelling, 0.5 % for seed and 3 % for fodder. The enforcement of such a threshold clearly requires quantitative detection systems such as quantitative competitive polymerase chain reaction (QC-PCR).

A large-scale certification program must be based on the testing of raw materials rather than finished products. It would be economically impossible for manufacturers to work with a programme that failed to indicate the GMO status of ingredients until their products were finished. In addition, some finished products have been so highly refined that they cannot be tested for GMO content, even by the most sensitive DNA tests.

The suitability of QC-PCR was successfully ring tested in Switzerland in 1998 by 12 analytical laboratories (1). The QC-PCR detection system was assessed by four independent determinations of the GMO content of certified reference material containing 0.5 % and 2 % Roundup Ready Soybeans (RRS), respectively. The relative standard errors of the means were 9 % and 2 % respectively.

However, it is important to recognise the difference between the potential and real sensitivity of the PCR method. PCR-based assays require additional steps that profoundly influence the reliability and sensitivity of the method as a whole. In particular sampling procedures and the procedures used to prepare the analytical

sample from the field sample are key to reliable and informative analysis of foods and agricultural products for GMO content. Yet this aspect of analysis is often neglected when laboratories develop GMO analytical services. The person who is paying for the analysis is usually seeking information about the composition of a "target", a large body of material, say cereals or pulses in a container or in a silo. In contrast, the analyst's result, with its associated uncertainty u_{an} refers to the subsample, the usually much smaller amount of material that is subjected to the analytical procedure. These two things are not identical. At most everything that is worth analysing is actually or potentially heterogeneous. Consequently, any sample is likely to have a composition that is different from the mean composition of the target, and no two samples will have the same composition (2). This variation in composition, even among properly collected samples, is quantified as the sampling uncertainty u_{sam}. The failure accurately to inform the end-user arises because the sampling variation is normally not taken into account when an uncertainty budget is assembled. As the squares of the uncertainties are cumulative, that is $u^2_{tot} = u^2_{sam} + u^2_{an}$, potentially serious misunderstandings could occur if the sampling uncertainty would be greater than the analytical uncertainty.

Let L be a lot and a_L the true unknown proportion of A in L from which increments and samples are selected (i.e. soy beans, maize kernels, wheat, flower, etc.) and which is to be estimated by analysis. The true unknown proportion of A in the analysed sample S will be designated as a_S.

Three issues need to be considered:

1. The sample submitted by the customer (the field sample) must be representative of the material from which it was taken.
2. The sample that is analysed in the laboratory (the analytical sample) must be representative of the sample submitted for analysis by the customer.
3. Laboratories that carry out chemical analysis of the composition of test materials have to be compared for their performance.

The primary sampling stage (sample S_1) takes place outside the laboratory. Sampling responsibilities are often undefined and the sampling qualification non-existant. A relative sampling error $TE_1 = (a_{S_1} - a_L)/a_L$ up to 1000 % cannot be excluded. The analytical laboratory is strictly responsible for assuring that the analytical sample at the second sampling stage (sample S_2) is representative of the field sample. A relative sampling error $TE_2 = (a_{S_2} - a_{S_1})/a_L$ up to 50 % is not impossible if the qualification of the analyst is insufficient. It can be assumed that the analysis is usually carried out by excellent analysts and the relative analytical error $AE = (a_R - a_{S_2})/a_L$ is small in most cases. The analytical result a_R is an estimate of a_{S_2}. The overall estimation error OE is the sum of three independent random variables: $OE = TE_1 + TE_2 + AE$.

As a consequence sampling should benefit from the same care and the same investments as analysis. The effort of the analyst in the laboratory is futile if sampling has not been carried out correctly. Analytical reliability is today limited not by

the analysts intrinsic qualities but by the lack of reliability of the samples submitted to the analytical process. Quality estimation is a chain and sampling is by far, its weakest link.

It must be noted that this document serves rather as a general starting point for the development of effective sampling strategies and sampling plans for GMO analysis and does not offer solutions to complicated and specific sampling problems. In this case specific sampling references as *Pierre Gy's sampling theory and sampling practice* (3, 4), *Cochran* (5) or a qualified statistician may be of additional help and guidance. For a strategy to quantify the analytical uncertainty it is referred to *Lischer* (6).

Definitions of representativeness, probabilistic sampling, correct sampling, etc., are given in the next section and a condensed summary of Pierre Gy's sampling theory will be presented.

The concept of heterogeneity

Definitions and notations

Sampling

The selection of a certain subset of constitutive elements making up the lot.

Constitutive elements

Units that are unbreakable and unalterable in the physical, chemical and mechanical conditions that prevail during sampling.

Increment

A group of particles extracted from the lot, in a single operation of the sampling device. It is important to make a distinction between an increment and a sample which is obtained by the reunion of increments.

Comminution

A crushing, grinding or pulverising stage that diminishes the fragment size of a lot, sample, or increment.

A set of units is said to be *homogeneous* when all units are strictly identical with one another. It is said *heterogeneous* when this condition is not strictly fulfilled.

Probabilistic sample

All constitutive elements making up the lot have a non-zero probability of being selected.

Correct sampling

All constitutive elements of the lot are given an equal probability of being selected and the increment and sample integrity is duly respected.

Relative sampling error

$e = (a_S - a_L)/a_L$, where a_S and a_L are the true critical contents of sample S and lot L. The random variable e can be characterised by its distribution and the values of its moments (mean or bias, variance, mean-square).

A resulting sample is said to be (practically) *accurate* when the expected value $m(e)$ of e is practically zero: $|m(e)| \leq m_0$, where m_0 is the maximum acceptable bias.

It is said to be *biased* if $|m(e)| > m_0$.

It is said to be (sufficiently) *reproducible* when the variance $s^2(e) \leq s_0^2$ (where s_0^2 is the maximum acceptable variance.

It is said to be *insufficiently reproducible* if $s^2(e) > s_0^2$.

A selection is said to be *representative* (i.e. accurate and reproducible) when mean-square $r^2(e) \equiv m^2(e) + s^2(e) \leq r_0^2 \equiv m_0^2 + s_0^2$

A sampling problem is considered *solvable* if it is possible to develop and implement a sampling plan characterised by an acceptable degree of representativeness that can be achieved at an acceptable cost. Theoretically all sampling problems are solvable; however, in a great number of cases, the notion of solvability is closely related to the notion of cost effectiveness.

A lot of particulate material is always affected by a certain amount of *heterogeneity*. The more heterogeneous the material, the more difficult the sampling operation. Before deciding on a corresponding sampling operation the amount of heterogeneity intrinsic to a given material has to be measured. The independent analysis of heterogeneity is a fundamental step since it provides information which goes far beyond the sole purpose of sampling. In a simplistic way the concept of heterogeneity can be described as a scalar or a function, the homogeneity of which is zero. Therefore, homogeneity is a limit case. The hypothesis made when it is assumed that a material is homogeneous is very dangerous because it allows anyone to solve all sampling problems associated with heterogeneity by oversimplifying them.

Classification of lots

From a theoretical standpoint a lot has always three dimension; however, in practice, one or even two of these dimensions can often be regarded of secondary importance. The fewer the dimensions, the easier is the solution of the sampling problem. We can encounter the following cases:

- Three-dimensional lots: The content of a ship, truck, railroad car, bag, jar etc., as long as one of these three-dimensional object is considered as the whole lot.
- Two-dimensional lots: A three-dimensional lot in which the thickness becomes negligible because it is very small compared to the two others dimensions.

- One-dimensional lots: Continuous and elongated piles, material on conveyer belts, streams, etc., or series of truck, railroad car, bag, jar etc., as long as these objects are considered as a set of non-random, discontinuous objects making up the lot, the order of which is highly relevant (chronological series).
- Zero-dimensional lots: The content of a series of truck, railroad car, bag, jar etc., as long as these objects are considered as a set of random, discontinuous objects making up the lot.

A zero-dimensional lot can be regarded as a suitable convention to describe a set of unarranged units. It can also be a one-dimensional lot for which the chronological order of various units has been lost. It is always possible to transform a three- or two-dimensional lot into a one-dimensional lot, in fact, it is done all the time in order to facilitate handling, transportation, reclaiming, homogenisation and so on.

A lot may be represented by either a *continuous* or a *discontinuous* model, depending on whether the observer looks at it from a distance or under a magnifying lens. The scale of the heterogeneity may dictate the kind of model it will be considered:

- As a discrete and discontinuous set if the main interest is the amount of heterogeneity introduced by the various fragments.
- As a continuous set, such as a flowing stream of material on a conveyor belt. The relevant dimension is time or distance and the main interest are its long-range and periodic heterogeneity fluctuations.

Many misconceptions in sampling may have their origin in the confusion between random populations where neighbouring fragments are totally independent from one another and time or space series where neighbouring fragments are statistically correlated. Exactly as homogeneity is the zero of heterogeneity, a random population is the zero of a chronological series. Experience shows that perfect disorder is the exception, and order or partial order is the rule. This originates in the fact that industrial activities are well framed in time and space, generating a correlation over time and space, and also in the fact that gravity is omnipresent, generating segregation along a vertical axis during transportation or handling of particulate materials.

Heterogeneity of a zero-dimensional lot

Let L be a lot of particulate material of mass M_L and critical content a_L composed of constitutive elements (fragments) F_i of mass M_i and critical content a_i; $i = 1, 2, \ldots, N_F$.

The heterogeneity h_i carried by fragment F_i within the lot is defined as:

$$h_i = N_F \frac{a_i - a_L}{a_L} \cdot \frac{M_i}{M_L}$$

Its variance is called Constitution Heterogeneity CH_L:

$$CH_L = \frac{1}{N_F}\sum_i h_i^2 = N_F \sum_i \frac{(a_i - a_L)^2}{a_L^2} \cdot \frac{M_i^2}{M_L^2}$$

When all fragments have the same mass the constitution heterogeneity CH_L is an estimator of the relative variance $(1\text{-}a_L)/a_L$ of the content a_i of the fragments. The constitution heterogeneity CH_L is an intrinsic property of the lot L (in a given state of comminution for particulate solids). Nothing can alter it. In particular, homogenisation has no effect on it.

As N_F is usually very large CH_L is not easy to calculate in most real cases. In practice the Heterogeneity Invariant IH_L is used, defined as:

$$IH_L = CH_L \cdot M_{m^*}$$

where $M_{m^*} = \frac{M_L}{N_F}$ designates the mass of the average fragment.

The lot L is now considered as a set of groups of adjoining constitutive elements G_n of mass M_n and critical content a_n; $n = 1, 2, \ldots, N_G$; such as increments taken by a sampler. The heterogeneity h_n carried by a single group of fragments G_n in the (zero-dimensional) lot L, is analogously defined as:

$$h_n = N_G \frac{a_n - a_L}{a_L} \cdot \frac{M_n}{M_L}$$

Its variance is called Distribution Heterogeneity DH_L:

$$DH_L = \frac{1}{N_G}\sum_n h_n^2 = N_G \sum_i \frac{(a_n - a_L)^2}{a_L^2} \cdot \frac{M_n^2}{M_L^2}$$

The distribution heterogeneity DH_L, however, can be modified; either it can be diminished by homogenisation or mixing, or it can be increased by promoting segregation.

Heterogeneity of a one-dimensional lot

Let L be a one-dimensional lot made of N_u discrete units of mass M_m (e.g. increments of a flowing stream, railroad cars, trucks, bags, etc.) arranged in chronological order and U_m one of these units with $m = 1, 2, \ldots, N_u$. Emphasis is placed not on the heterogeneity fluctuation within each unit which could be treated as a zero-dimensional lot but on the heterogeneity fluctuations between units. As already defined for a zero-dimensional lot, the total heterogeneity carried by the unit U_m of a one-dimensional lot can be expressed as follows:

$$h_m = N_u \frac{a_m - a_L}{a_L} \cdot \frac{M_m}{M_L}$$

The study of a large number of chronological series leads to the conclusion that in many cases they are the result of the accumulation, around the average value a_L, of three kinds of fluctuations with independent causes:

1. A short-range term, h_{1m}, mainly random, discontinuous at every instant, reflecting the random nature of constitution heterogeneity.
2. A long-range term, h_{2m}, mainly continuous, representing trends between units.
3. A cyclic term, h_{3m}, continuous, such as cycles introduced by reclaiming operations.

The heterogeneity h_m can be written as $h_m = h_{1m} + h_{2m} + h_{3m}$.

In order to determine the variances of the components h_{1m}, h_{2m} and h_{3m} the *variogram V(j)*:

$$V(j) = \frac{1}{2(N_u - j)} \sum_m [h_{m+j} - h_m]^2, j = 1, 2, \ldots .$$

has to be calculated. Variograms have been successfully used in mining and quality control to predict sampling variances.

It is possible to characterise a chronological series in two different ways:

- An overall characterisation of the heterogeneity carried by the statistical population of the h_m using a scalar such as the variance. The order of the units, even if known, is voluntarily considered as irrelevant.
- A sequential characterisation of the heterogeneity carried by the chronological series of h_m using a function such as the variogram. The order of units described by the values of the subscript m is highly relevant.

Error components

Sampling is a stepwise process. At each sampling stage several different independent error components have to be taken into account (3).

The short-range heterogeneity fluctuation error CE_1

A sample S selected from a (zero-dimensional) lot L is affected by an error specifically related to the constitution heterogeneity CH_L of the same lot. The fundamental error FE is the minimum error generated when collection a sample of a given mass. The minimum is reached only under one statistical condition – fragments making the sample shall be collected strictly at random, and one by one. Of course, it does not happen this way in practice. When collecting an increment to make up a sample, this increment is likely to be made of many fragments. Then statistically speaking, one sample is not made of strictly random fragments, but only of random groups of fragments. Consequently, an additional error will be introduced and the larger the group, the larger the error. The error introduced by distribution heterogeneity DH_L is called segregation and grouping error GE. The short-range heterogeneity fluctuation error is the sum of the fundamental error and the segregation and grouping error: $CE_1 = FE + GE$.

The continuous sampling selection error CE

Industrial activities are characterised by a constant need to transport materials (ores, concentrates, coal, cereals, chemical, etc.) from one location to an other. The practical implementation of such activities necessarily generates long piles, running materials on conveyor belts, and streams that are defined as one-dimensional lots. It is always possible to implement a correct sampling operation on one-dimensional lots. Obviously what has been said of a zero-dimensional lot is still true for a one-dimensional lot which will be affected by a certain constitution heterogeneity coupled with a transient term which is the distribution heterogeneity; however, a one-dimensional lot is nearly always generated by chronological operations. Consequently it will be affected by fluctuations that are mainly reflecting human activities at the mine, at the mill, at the processing or chemical plant, etc. These are not intrinsic properties of the material making up the lot – these are trends and they lead to a new concept of heterogeneity that can be subdivided into two terms:

1. The heterogeneity h_2 introduced by long-range trends, which could be defined as a large scale segregation and
2. the heterogeneity h_3 introduced by cyclic phenomena. The errors introduced by these types of heterogeneity are the long-range heterogeneity fluctuation error CE_2 and the periodic heterogeneity fluctuation error CE_3.

Therefore, in the case of a chronological series of units, the heterogeneity introduced by random constitution heterogeneity is defined as the small scale heterogeneity h_1, which introduces an error defined as the short-range heterogeneity fluctuation error CE_1. The total heterogeneity h can be expressed as $h = h_1 + h_2 + h_3$ and the continuous selection error as $CE = CE_1 + CE_2 + CE_3$. CE_1 will serve as a link between the continuous model and the discrete model. If a one-dimensional lot was considered as a zero-dimensional lot, h_3 would cancel and h_2 would become part of h_1, which is obvious since the lot would be considered as a random population.

The increment materialisation error ME

Thus far, the lot was considered as a one-dimensional continuous object and the increment selection was based on imaginary points within the domain of interest, but the real points are made of fragments or groups of fragments and the discrete nature of these units should be taken into account. The same reasoning can be made with the splitting process of a zero-dimensional lot. The materialisation is achieved by first implementing a correct increment delimitation, then a correct increment extraction, which are error-generating processes. The increment materialisation error ME is defined as the sum of the delimitation error DE and the extraction error EE: $ME = DE + EE$.

The total sampling error SE

In a proper sense the sampling error SE is introduced by the bulk reduction of a lot after the selection of a series of increments, the materialisation of which is mak-

ing up the sample. Sampling is considered as a combination of only two categories of operations, an immaterial selection process which leads to the continuous selection error *CE*, and a materialisation process which leads to the materialisation error *ME*: $SE = CE + ME$. This sum is defined as the sampling error generated by only one stage.

The overall estimation error OE

In addition to the selection process in general all non-selecting operations carried out on the lot and on the successive samples generated by various sampling stages have to be taken into account. These non-selecting operations, or at least some of them, are likely to be present between each sampling stage, and are defined as *preparation stage*. A preparation stage is an error-generating process which may consist of transfer, comminution, screening, mixing, drying, filtering, weighing, packing, etc. The generated error, usually an accidental error, is defined as the preparation error *PE* and the total sampling error *TE* as the sum of the sampling error *SE* and the preparation error *PE*, generated for each sampling and corresponding preparation stage: $TE = SE + PE$. If there are *N* sampling and preparation stages, *N* total sampling errors will be generated, and if *AE* is the analytical error the overall estimation error is defined as:

$$OE = AE + \sum_{n=l}^{N} (SE_n + PE_n)$$

Probabilistic and non-probabilistic sampling processes

Examples of incorrect selection processes

Grab sampling

Philosophy: *Catch whatever you can in the cheapest possible way!*

Grab sampling cannot be accurate because some units making up the lot have a zero probability of being selected. Furthermore, grab sampling cannot be accurate because the selecting probability between units making up the lot cannot kept constant.

Purposive sampling

The operator chooses the fragments he regards as "representative of the material". From a statistical point of view it is difficult to admit that the choice is probabilistic, even if he or she is creditable and honest ("sworn sampler"). Purposive sampling cannot be accurate and is very likely to be inequitable as well.

Sampling with thief probes and auger

These are only improved grab sampling techniques. The idea is to extract a column representing the entire thickness of the lot at a preselected point. These techniques are probabilistic only when the place to perform the extraction is also selected in a probabilistic manner which is rarely done. For example, drilling at the centre of a barrel, following the diagonal of a bag, drilling at pre-established points from a truck to the bottom, drilling on a waste pile, and so on. Assuming that these techniques could be probabilistic is a very optimistic assumption. Experimental studies prove that they are scarcely correct. Sampling with thief probes and auger cannot be accurate because of serious delimitation and extraction problems encountered during penetration.

Common properties of non-probabilistic selection processes

1. They are implicitly based on the unrealistic hypothesis of homogeneity.
2. An important fraction of the lot is submitted to the sampling process with a zero probability of being selected. This is a critical point for sampling equity, especially during commercial sampling.
3. There is no possible theoretical approach. It is impossible to logically connect the various sampling errors to the mode of selection.
4. They generate uncontrollable biases $|m(e)| >> m_0$ and unacceptable variances $s^2(e)| >> s_0^2$.

Probabilistic sampling of movable lots

A batch of particulate material is said to be *movable* when it is small or valuable enough to be handled in totality for the sole purpose of its sampling. One of the two following probabilistic processes can be used to sample movable lots:

1. The increment process: The batch of material is transformed into a one-dimensional stream, then a cross-stream sampler collects a certain number of increments to make up the sample.
2. The splitting process: The batch of material is partitioned into several fractions, one of which is selected at random as a sample.

Analysis of the increment sampling process

A good example of the increment sampling process is the sampling of a flowing stream:

- at the discharge of a conveyer belt;
- across a conveyor belt;
- at the discharge of a pipe or a hose;
- across a river, and so on

by means of a correctly designed cross-stream sampler.

There are three ways of reducing the flowing mass of a stream:

1. Taking the whole of the stream during a fraction of the time.
2. Taking a fraction of the stream during the whole of the time.
3. Taking a fraction of the stream during a fraction of the time.

2 and 3 are never probabilistic. They are structurally biased. Only cross-stream sampling 1 is probabilistic. It can easily rendered correct and therefore accurate.

Let L be a lot flowing from time $t = t_0$ to $t = t_n$. The critical content of analyte A at time t is designated by $a(t)$. Increments to form the sample S are taken at time $t = t_i$, $i = 0,1, \ldots, n-1$. The relative continuous selection error $CE = (a_S - a_L)/a_L$ of sample S can be approximated by the difference of the integrals of the function $a(t)$ and of the step-function $a(t_i)$, $i = 0,1, \ldots, n-1$. The true content a_L can be expressed as

$$a_L = \frac{1}{t_n - t_0} \int_{t_0}^{t_n} a(t)dt$$

and the integral a_S of the step function $a(t_i)$ by

$$a_S = \frac{1}{t_n - t_0} \sum_{i=0}^{n-1} a(t_i)(t_{i+1} - t_i) \; .$$

On a large scale, the stream can also be made of large units such as trucks, railroad car loads, sacs, barrels, or jars arranged in a chronological order as a one-dimensional lot. Each unit becomes one potential increment. The increment sampling process can be broken up into a sequence of four elementary and independent steps:

1. The sampling point selection: All points along the one-dimensional lot are submitted to a selection scheme that can be:
   - either systematic with a random starting point;
   - either stratified random;
   - or random.
2. The increment delimitation: Moving through the lot a point is selected, then the sampling device delimits the geometrical boundaries of the domain from where the *extended increment* should be extracted. The extended increment is a volume that does not take into account the particulate nature of the material.
3. The increment extraction: Now it is necessary to take the particulate nature of the material into account. The sampler extracts a certain number of fragments making up the *fragmental increment*. The fragmental increment must coincide with the set of fragments whose centre of gravity falls within the boundaries of the extended increment.
4. The increment reunion: The set of fragmental increments is called the fragmental sample.

Analysis of the splitting process

Typical examples of the splitting process are

- Coning and quartering;
- Alternate or fractional shovelling;
- Riffling and so on.

The splitting process can be broken in a sequence of four elementary and independent steps presenting a great similarity with the sequence observed with the increment sampling process:

1. The fraction delimitation: The sampling device delimits the geometrical boundaries of the domain occupied by the geometrical fractions of the lot. Three different cases can be encountered:
   - Stationary lot, moving tool: coning and quartering, alternate shovelling;
   - moving lot, stationary device: riffle divider, revolving feeder – sectorial splitter;
   - moving lot, moving device: stationary feeder – sectorial splitter.

   As for the increment sampling process, the geometrical delimitation does not take into account the particulate nature of the material making up the lot.
2. Separation of fractions: This operation takes the particulate nature of the material into consideration. The fragmental fraction must, more or less, coincide with the set of fragments whose centre of gravity falls within the boundaries of the geometrical fractions.
3. Reunion of fractions: Fractions are regrouped together according to a systematic scheme in order to provide a given set of potential samples. Of course the set of these samples is the entire lot L.
4. Sample selection: This selection must be probabilistic; therefore the real sample or samples shall be selected at random.

Comparison of the increment process with the splitting process

From a logical and practical standpoint there is a fundamental difference between both processes:

- For the increment process, the selection is made before delimitation and extraction steps.
- For the splitting process, the selection is made after the extraction step.

As a consequence, the splitting process can be equitable even when it is technically biased; however, the increment process is equitable if, and only if, it is technically unbiased. This is an important detail that can go a long way in commercial sampling. Even if potential samples created by splitting are systematic different, the actual sample is chosen at random. The expected error over the long run is zero. This is not the case for the increment process because selection precedes the materialisation of the sample.

Good sampling strategy

The principal objective of any sampling process is to provide a representative sample, the true unknown content of which is noted a_S. The estimate a'_S of a_S should provide an unbiased and precise estimator of the unknown content a_L of the lot L. Theory of sampling shows that a correct sampling process is always accurate. Sampling correctness does not depend on the properties of the material to be sampled. Correctness is an intrinsic property of the sampling process as long as the integrity of the equipment is not damaged. The conditions of correct sampling involve properties of the sampling device or method such as design, construction, lay out, operation and maintenance. Control of the correctness of a sampling station by a specialist consists of performing a critical inspection and a few very simple measurements.

A good sampling strategy should provide the following chronology:

1. Study of the heterogeneity of the material of a given lot, either a zero-dimensional lot or a one-dimensional lot.
2. Optimisation of the sampling protocols to minimise fundamental error *FE*, grouping and segregation error *GE*, long-range-heterogeneity fluctuation error CE_2, and periodic heterogeneity fluctuation error CE_3.
3. Control of sampling correctness (i.e. choice of the sampling equipment) in order to eliminate the increment delimitation error *DE*, the increment extraction error *EE* and the preparation error *PE*.

The standardisation of a sampling strategy and of a correct sampler can be very general, while the standardisation of a sampling protocol can only be local. Each case is unique and the statistician alone cannot decide upon an appropriate sampling plan. An effective coordination between those who have knowledge in the domain in which the problem takes place and the statistician is a must.

Factors determining the establishment of sampling plans for GMO analysis

After the statistical concepts have been understood, both general and specific sampling objectives must be defined prior to any sampling activities. A number of factors need to be taken into consideration in the development and adoption of appropriate sampling plans for GMOs in raw materials (7). One critical factor will be the threshold limit which is set for acceptance of the presence of GM material – the lower the limit the greater the demands will be upon the sampling plan. It is also essential that a quantitative method of analysis is available which has a better sensitivity than the threshold and an adequate precision. The speed requirement for making a decision and the point at which the sample can be taken will also be important factors which will be taken into consideration in proposing a sample regime. The ideal situation is when material such as soy beans or maize kernels are being unloaded from barges or trucks where there is the option of continuous sampling.

Finally in deciding upon a sampling plan to be adopted there is a need to decide upon the level of an acceptable producer risk and conversely the level of acceptable consumer risk.

Conclusions

- Sampling should benefit from the same care and the same investments as analysis. The sampling and analytical errors, biases and variances are additive.
- Theory of sampling shows that a correct sampling process is always accurate. Sampling correctness does not depend on the properties of the material to be sampled.
- Correct sampling is often completely uncorrelated with sampling costs. Excuses to perform incorrect sampling cannot be justified by time and money limitations.
- Analysts should refuse to give results whenever they are not satisfied that the samples they receive are representative.

Acknowledgment

This work has been supported by the Swiss Federal Office of Public Health, Contract No. 00.000411 (Reg. 2.26.01.-39)

Summary

One of the most important and often neglected elements essential for accurate genetic testing and certification is sampling. Testing results only reflect the amount of genetically modified organisms in the sample received in the laboratory.

Assurance that a laboratory sample is representative of the material from which it is taken is provided by correct sampling and a sample size large enough to allow analysis to the desired precision.

A condensed summary of Pierre Gy's sampling theory and sampling practice is presented. A backbone of the theory is the study of the various forms of heterogeneity. It is shown that only splitting and cross-sampling from a flowing stream provide representative samples.

Zusammenfassung

Eine schlecht geplante oder unsorgfältig durchgeführte Probenahme lässt sich auch durch eine qualitativ hochstehende Laboranalyse nicht mehr korrigieren. Probenahme und Analysis erfordern deshalb die gleiche Aufmerksamkeit und Sorgfalt.

Das Ziel aller Stichproben-Auswahlverfahren besteht darin, eine Teilprobe auszuwählen, deren Zusammensetzung mit der gesamten zu untersuchenden Materialmenge übereinstimmt. Dies wird durch eine korrekte Probenahme erreicht. Je nach gewünschter Genauigkeit ist eine mehr oder weniger umfangreiche Stichprobe erforderlich.

Die Anwendung von Pierre Gy's Theorie und Praxis der Probenahme macht es möglich, repräsentative Stichproben zu gewinnen. Es wird gezeigt, dass repräsentative Teilproben aus umfangreichen Materialmengen nur mit einer mechanischen Probenahme mit geeigneten Geräten aus in Bewegung befindlichen Materialmengen gewonnen werden können. Probenteiler können nur bei kleinen Materialmengen verwendet werden

Résumé

Échantillonnage et analyse chimique doivent être exécutés avec les mêmes soins. Un mauvais échantillonnage peut ôter toute signification aux résultats d'analyse.

Seul un échantillonnage correct et une taille d'échantillon suffisamment grande fournissent des échantillons représentatifs et assurent une analyse avec une précision donnée.

Les méthodes d'échantillonnage établies par Pierre Gy permettent la prise d'échantillons représentatifs. Il est montré que seul l'échantillonnage par fractions et des coupes transversales de la matière en cours de mouvement sont garants d'échantillons représentatifs.

Key words

Correct sampling, Representativeness, Sampling uncertainty, Heterogeneity, Pierre Gy's sampling theory

References

1 *Hübner, P., Studer, E.* and *Lüthy, J.:* Quantitation of genetically modified organisms in food. Nature Biotechnology, **17**, 1137–1138 (1999).

2 *Thompson, M.:* Sampling the uncertainty that dares not speak its name. J. Environ. Monit., 1, 19N–21N (1999).

3 *Pitard, F.:* Pierre Gy's sampling theory and sampling practice. CRC Press, Boca Raton, Florida 1993.

4 *Gy, P.:* Sampling of particulate materials – Theory and practice, 2nd ed., Elsevier, Amsterdam 1982.

5 *Cochran, W.G.:* Sampling techniques, 3rd ed. John Wiley & Sons, New York 1977.

6 *Lischer, P.:* Good statistical practice in analytical chemistry. In: Grigelionis, B. et al. (eds), Prob. theory and math. Stat., Proceedings of the 7th Vilnius Conference, 469–80 (1999).

7 *Gilbert, J.:* Sampling of raw materials and processed food for the presence of GMOs. Food Control **10**, 363–365 (1999).

Corresponding author: Peter Lischer, Ahornweg 46, CH-3095 Spiegel bei Bern, E-mail: plischer@bluewin.ch