Communicating the importance of Representative Sampling

- to governments
- to the European Union Commission – and its agencies
- to management
- to regulators
- to enforcers of regulations
- to analytical chemists
- to statisticians
- to commodities trading communities

is critically important to insure implementation of effective sampling protocols capable of enforcing regulatory specifications in a fair manner.
The Responsibility - of Commissions, Governments, Regulators, Inspectors...

If companies, agencies, regulators, inspectors ... cannot see the value of correct, i.e. representative sampling - it is the responsibility of the relevant top-level authority to enforce proper standards through...

- Education of management and technical expertise - to acquire the knowledge base and to secure the appropriate resources,
- Training of key technical personnel, to get documentable results,
- Overview education of regulators and enforcers, to monitor and verify the quality of product shipments,
- Facilitate creation of a valid database for all relevant commodities - so that statisticians can perform reliable risk assessments ...
Acquisition of Reliable Competences for Authoritative Regulation

A correct, balanced strategy is needed:

The three-legged Table

- Fair Enforcement of Correctly Defined Regulations
- Emphasis on causes of problems by proactive management
- Capability to understand variability and to perform reliable statistical studies
- A strong commitment to good sampling and good laboratory practices

Francis Pitard's drawing protected by copyright law, 2004
However, implementing *correct* sampling is easier said than done.

Exactly like safety issues, it must...

- be **standardized** through:
  - correctness,
  - guidelines,
  - sustained training,
  - enforcement auditing.

- be **monitored** for its added value to the scientific foundations:
  - fair, achievable regulations,
  - improved conciliation between parties involved,
  - added value to the long-term public health.

Representative; “correct” sampling: Theory of Sampling (TOS), *Pierre Gy*
Regulators must learn to better understand variability:

- **Small-scale variability**, which can be viewed as Misleading Variability:
  It is a nuisance. This is termed $V[0]$ (TOS)

- **Large-scale variability**, which can be called the Leading Variability.
Small-scale Variability: 
The term $V[0]$ in a variogram

Four solutions to minimize a catastrophic inflation of $V[0]$:

- **Optimizing Sampling Protocols,**
- **Implementing Representative Sampling Protocols using correct sampling systems,**
- **Preserving sample integrity,**
- **Minimizing Analytical Error.**

These are the Critical Success Factors for reliable quantification of GMO levels
Two critically important TOS- issues:

- **Fundamental Sampling Error (FSE)**
  (e.g., sample and sub-samples mass)
  For trace constituents, beware of devastating Poisson processes that may take place when the variance of FSE is larger than +/-16%.

- **Grouping & Segregation Error (GSE)**
  (e.g., Homogenization and number of increments)
  There are plenty of sources for intermediate-scale inhomogeneities, which are the root cause of GSE
Practical Implementation of Sampling Protocols

The nightmare of practical sampling

Three major sources of sampling bias:

Invalidity of manual, grab-sampling:

- **Increment Delimitation Error IDE**
  
  (Every part of the lot to be sampled must have exactly the same chance of becoming part of the sample – the Fundamental Sampling Principle)

- **Increment Extraction Error IEE**
  
  (The sample recovery error: The sampling system must not be selective)

- **Weighting Error WE**
  
  (Sampling systems must be reasonably proportional.)
Another major source of *sampling bias*:

**Increment Preparation Errors (IPE)**

(Errors taking place between sampling stages, or after the sample is taken)

- Contamination
- Losses (e.g. dust fraction, moisture ..)
- Alteration
- Human errors, ignorance
- Fraud
Practical example: Sampling of *loading* a Cereal Shipment

There are critically important issues that are necessary to control in assessing the contents of any commodity shipment.

**These are the Sampling Errors:**

It is critical that all sampling errors can be identified, assessed and appropriately reduced/eliminated.
Identify the possible sampling errors (IDE, IEE, IWE, or IPE?) taking place at each of the 11 locations. **Name each and give solutions.**

Francis Pitard's drawing protected by copyright law, 2004
Practical Exercise: Rotating Vezin Sampler

A common secondary or tertiary sampler

Name the possible sampling errors (IDE, IEE, IWE, or IPE?) taking place at each of the locations. **Name them and give solutions.**

Francis Pitard's drawing protected by copyright law, 2004
Practical Example:
The Cross-belt Sampler

A popular, but very dangerous sampling system

Name the possible sampling errors (IDE, IEE, IWE, or IPE?) taking place at each of the following locations. **Name them and give solutions.**

Francis Pitard's drawing protected by copyright law, 2004
Why is it that proper TOS training essential?

Because all the possible problems created by each point addressed in the three examples above, that could have been solved within minutes, usually are the object of unnecessary agony ... in the form of:

- doubts and arguments,
- time-consuming meetings,
- endless arguments with manufacturers of sampling equipment,
- endless arguments with regulators and enforcers of regulations,
- very expensive bias tests followed by doubtful statistical analysis,
- unfortunate uncertainty, giving motives for unnecessary litigation.

Furthermore ...

because each issue can lead (- will lead) to devastating money losses for all parties involved because of faulty decisions and, not least important, through the inevitable litigation process.

ALL THIS COULD BE AVOIDED THROUGH PROPER UNDERSTANDING OF THE THEORY OF SAMPLING (TOS)
The Poisson Distribution often occurs when a trace constituent, made of rare isolated particles or clusters, appears in a given material.

It also occurs when the sample weight is too small by one or several orders of magnitude.

Let’s call $r$ the actual number of trace element particles or clusters in a sample.

$\theta$ the hypothetical average number of trace element particles or clusters per sample, or average of the distribution.

$P(x=r)$ the probability of $r$ trace element particles or clusters appearing in a sample.

$$P(x = r) = \frac{\theta^r}{r!} e^{-\theta}$$

Variance = $\theta = npq = np$ since $q$ is close to 1

Mean value = $\theta = np$
When samples taken contain the constituent of interest in discrete grains or clusters, and they are subsampled in such a way that the subsamples also contain discrete grains of reduced size, a double Poisson Distribution of the assay values is likely.

If each sample contains a limited number $Z$ of the constituent grains, and the subsamples they generate also contain a limited number $z$ of constituent grains, the distribution of assay values is doubly Poisson.

The probability $\pi$ of $r$ grains of the constituent appearing in any subsample is determined by the sum of the probabilities of $r$ grains being generated from samples with $n$ grains.

$$\pi_r = \sum \frac{Z^n e^{-n} n^r}{r! n!}$$

with $n = 0, 1, 2, 3, \ldots$
The Special Case of GMOs: Poisson Processes

- The argument that GM grain does not follow a Poisson distribution is not valid. It is not the shipment (cargo) that can be Poisson distributed; it is how we sample and sub-sample the shipment that is the problem. This is a critical difference!

- Attempts to adapt random Poisson’s mathematical properties to a non-random GM material distribution events is an exercise in futility.

- The point is: - if/when the Standard Deviation $S$ of the FSE is larger than ±16%, a Poisson Process will progressively take place, with its illusion making consequences.

CONCLUSION: Be preventive.

- Regulatory sampling, sub-sampling, and analytical protocols must make sure Poisson Processes do not enter the regulatory enforcement protocols, guidelines, procedures and regulations.
Summary:
Who are the enemies?

Remember this sign:

\[ \sum_{n} S^{2}_{\text{FSE}_{n}} \]

\[ \sum_{n} S^{2}_{\text{GSE}_{n}} \]

\[ \sum_{n} S^{2}_{\text{HES}_{1}} \]

\[ \sum_{n} S^{2}_{\text{IDE}_{n}} \]

\[ \sum_{n} S^{2}_{\text{IWE}_{n}} \]

\[ \sum_{n} S^{2}_{\text{AE}_{n}} \]

\[ \sum_{n} S^{2}_{\text{IEE}_{n}} \]

\[ \sum_{n} S^{2}_{\text{IPE}_{n}} \]

Francis Pitard's drawing protected by copyright law, 2008

Remember this sign:
Large-scale Variability: The variability that needs to be fairly assessed for any given shipment.

• The argument should not be if GM grains are randomly or non-randomly distributed in a particular shipment.

They are likely distributed both ways ...

• The argument should be: Is the random variability – artificially introduced by inferior protocols - small enough to ensure that it does not interfere with the fair assessment of the non-random variability within one shipment?
Large-scale Variability:
How to assess it in a fair way?

- The KeLDA studies shows that:
  - Random variability observed in shipments is significant, but only moderate
  - Non-random, structured large-scale variability is overwhelming, and therefore a critical issue for establishment of logical, reliable regulatory sampling protocols
  - Large-scale variability within shipments often shows cyclical behavior that must also be effectively addressed in the regulatory protocols.

Regarding KeLDA studies I and II, please see the following contribution by Esbensen, Paoletti & Minkkinen: Representative Sampling of Large Kernel Lots (GMO and trace constituents): Variographic Analysis & Total Error Estimation (TOS).
CONCLUSION:
Stratified Random Sampling is the key.

- Regulatory protocols must be **preventive**, so no substantial artificially introduced random variability contaminates the database created to assess the variability of one shipment (i.e., sample and sub-sample mass issues).

- Each shipment must be **divided into a certain number of lots** (e.g., 2500-ton lots) that must be assessed separately for their average GM content, in order for the large-scale non-random variability to be properly assessed.

- Each lot should be assessed for its **average content** by collecting a sufficient number of increments to integrate the lot variability in such a way that no hidden variability may escape the inspection process.

- Each increment must be collected at **random within a 100-ton stratum**, so cyclic behaviors do not generate a sampling bias.
RECOMMENDATIONS:
Definition of Data Quality Objectives

- It is not sufficient to define a maximum regulatory GMO content (e.g. 1.00%).
- GMO content must be determined in several lots from a single shipment (e.g., 2500-ton lots).
- Failure to comply with one lot is failure for the whole shipment for regulatory compliance.
- Variability between lots must be such that it is unlikely that a small fraction of any given lot would escape inspection. Proper statistical characterization of this event must be the responsibility of further study as delegated by the relevant authorities.
- Regulations must address sample mass and sampling correctness.
- Regulations must address size of lots.
- Regulations must address increment frequency within lots.
- Regulations must enforce random selection of increment within each stratum of any given lot.
THE FINAL CARDINAL RULES

1. The **elimination** of non-valid, **non-probabilistic** grab-sampling is mandatory.

2. **Reliance** on optimized, proper **TOS**-compliant sampling protocols is mandatory.

3. **Over all**: **Reliance on correct sampling is the singular Critical Success Factor.**

Non-compliance to any of these rules **is acceptance of unscientific and unethical behavior.**