Uncertainty from sampling food—an empirical approach

Michael Thompson
Birkbeck University of London
Malet Street, London WC1E 7HX
m.thompson@bbk.ac.uk

Michael H Ramsey
University of Sussex
Brighton BN1 9QG
m.h.ramsey@sussex.ac.uk

Scientists involved in this project

From the University of Sussex:

Prof Michael Ramsey
Dr Bas Geelhoed
Dr Jennifer Lyn
Ilaria Palestra

From Birkbeck University of London:

Prof Michael Thompson

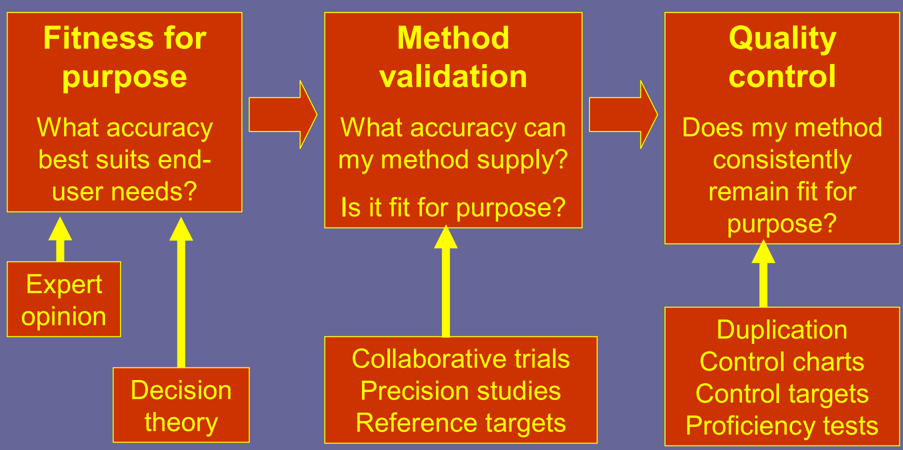
From the Food Standards Agency:

Dr Roger Wood
Dr Andy Damant
Dr Ellis Evans

‘Quality’ concepts and actions

- *Fitness for purpose*: what uncertainty is best for the customer?
- *Method validation*: can the method produce a suitably low uncertainty?
- *Internal quality control*: have things changed since validation? (*i.e.*, did the method work well on the day?
- *Proficiency testing*: does the whole system really work?

How it all fits together



The traditional approach—sampling considered separately from measurement.

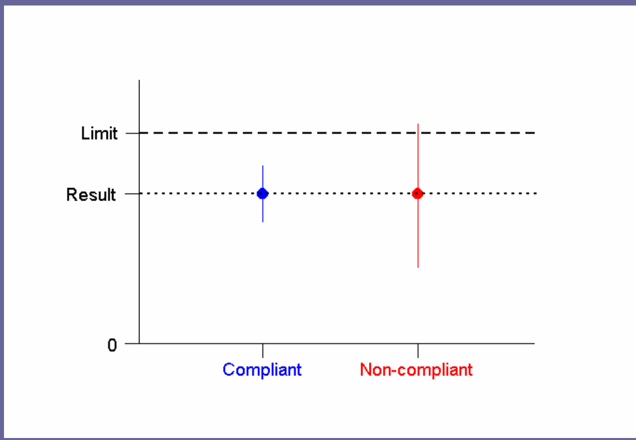
- Design ‘correct’ sampling protocol to give a ‘representative’ sample.
- Train sampler to apply the protocol.
- Assume $u_{\mathrm{sam}} = 0$.

The traditional approach is logically untenable. Why?

- Customers (and other stakeholders) need to know the total combined uncertainty to make informed decisions *about the target*.

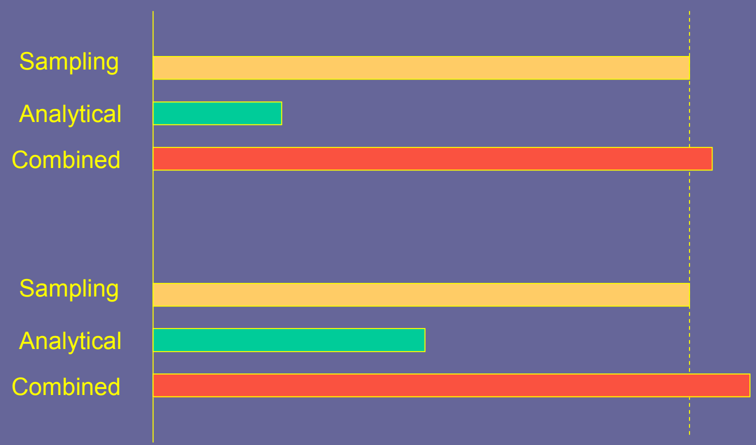
$$u = \sqrt{u_{sam}^2 + u_{an}^2}$$

Uncertainty in compliance decisions



Combining uncertainties

$$u = \sqrt{u_{sam}^2 + u_{an}^2}$$



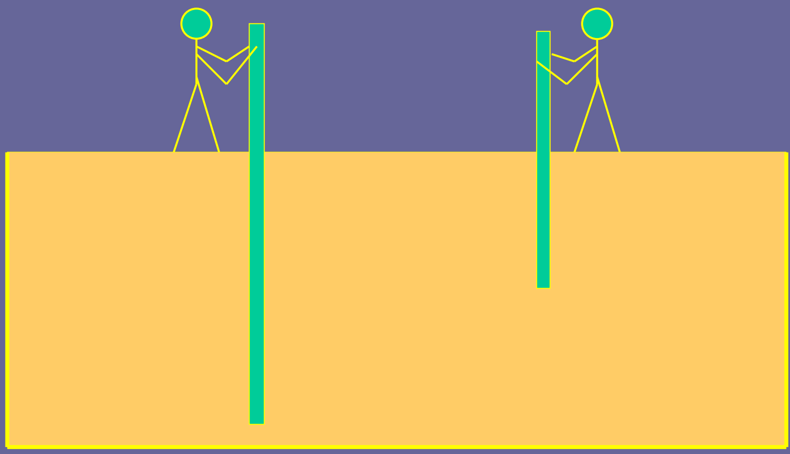
Components of sampling uncertainty

- *Bias*—difficult, often impracticable to address.
- *Precision*—Easy to address so long as a random element can be introduced into replicating the procedure.

Sampling bias

- Some experts think that sampling bias does not exist.
- Essentially they hold that sampling methods are empirical, *i.e.*, give an unbiased sample by definition.
- That is not generally correct—it is easy to see how sampling bias could arise in practice.

One way of taking a biased sample!



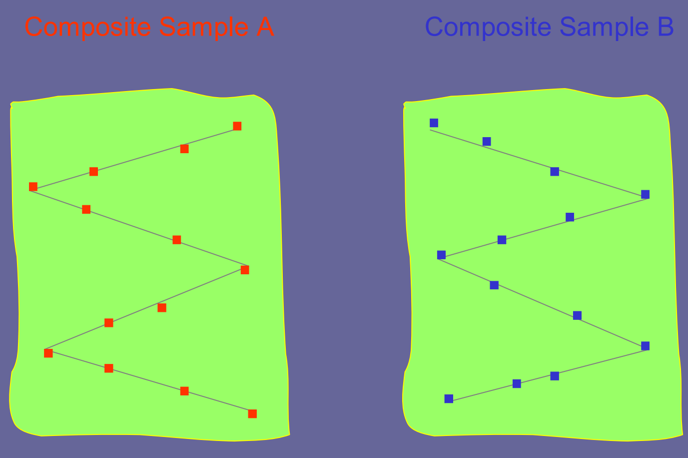
Addressing bias

Analytical	Sampling
Reference material	Reference target (Severe problems with cost, stability)
Reference method *vs.* candidate method (multiple test materials)	Reference method *vs.* candidate method (multiple test targets)

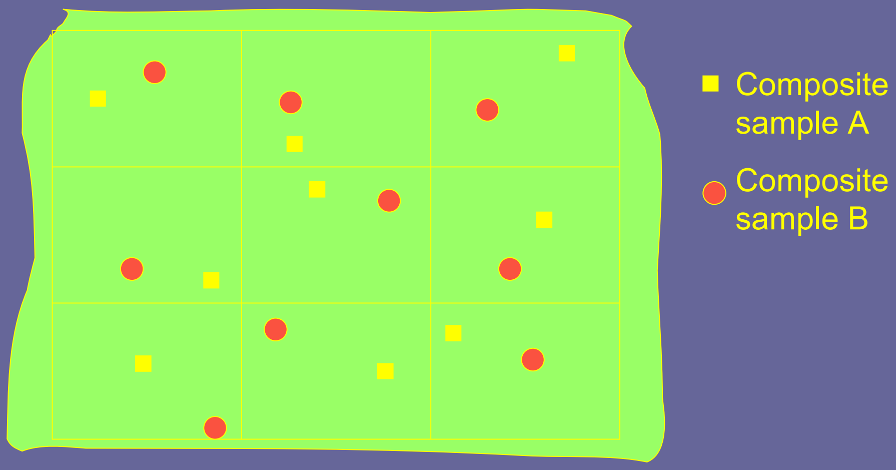
Sampling precision

- Variations in execution of procedure.
- Variations in composition (heterogeneity) of target.
  - Sampling precision may vary from target to target of the same nominal type.
  - Initial validation of the sampling protocol needs to be supported by ongoing checks (internal quality control).
- Good estimation of precision needs *RANDOM* replication of sampling.

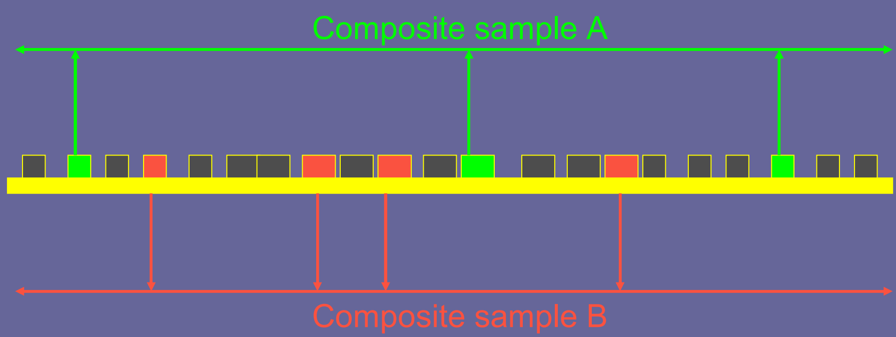
Random duplication—sampling to a pattern



Stratified random design



Sampling from a conveyor belt



Collaborative trial

- Requires:
  multiple targets, multiple samplers, duplicate samples, duplicate analysis (random repeatability conditions).
- Provides:
  analytical repeatability variance,
  between-sample (repeatability) variance,
  between-SAMPLER (reproducibility) variance.
- Drawbacks:
  VERY expensive.
- Current usage:
  research only.

Method validation—nitrate in lettuce

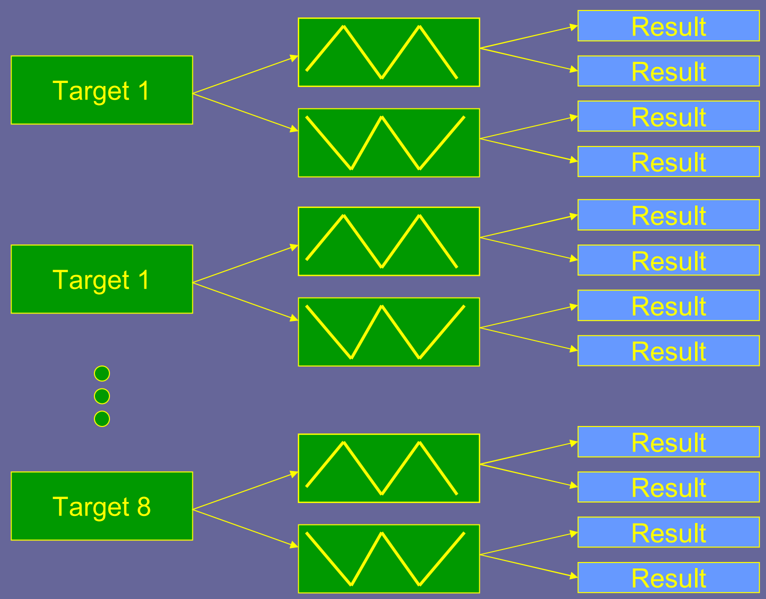
- Nitrate a potential risk to human health
- EU threshold is 4500 mg kg^{-1} for batch concentration
- Current sampling protocol specifies taking 10 heads to make a single composite sample from each batch.
- Sampling uncertainty unknown

Sampling of lettuce from one bay of greenhouse

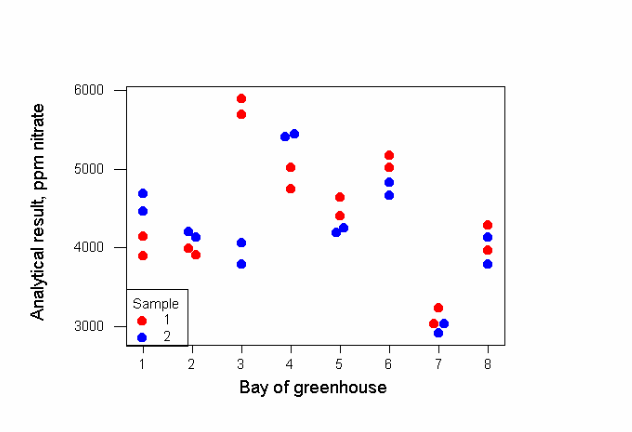
Randomisation is important…

…but not always exactly feasible.
Here we use systematic replication.

Balanced sampling design



Results of the balanced experiment



Statistics from robust ANOVA

$$\hat{\sigma}_{anal} = 168$$

$$\hat{\sigma}_{samp} = 319$$

$$\hat{\sigma}_{comb} = 361$$

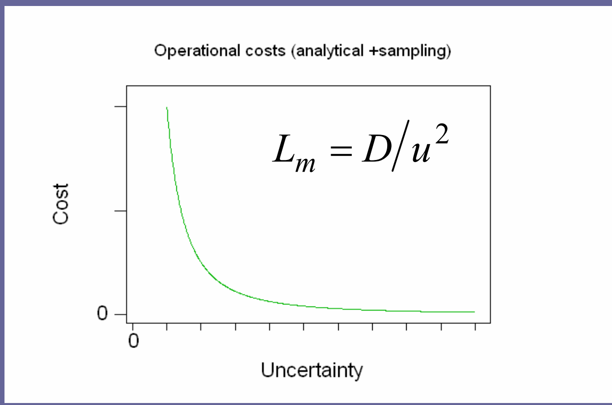
$$\overline{\hat{\mu}} = 4346$$

•Is the accuracy fit for purpose?
•(*Note*: σ is equivalent to standard uncertainty if measurement bias is absent.)

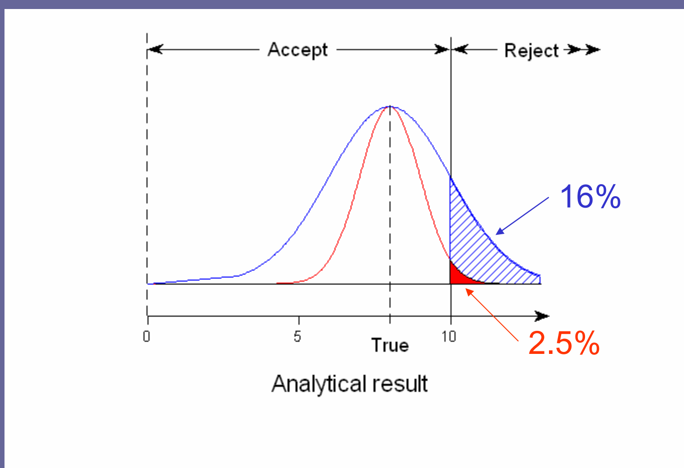
Fitness for purpose

- A result is fit for purpose when it maximises its expected utility.
- This means roughly that we need to minimise expected costs in the long term.
- There are operational costs of sampling and analysis.
- There are potential costs resulting from incorrect decisions based on the result.
- Both of these costs depend on uncertainty.

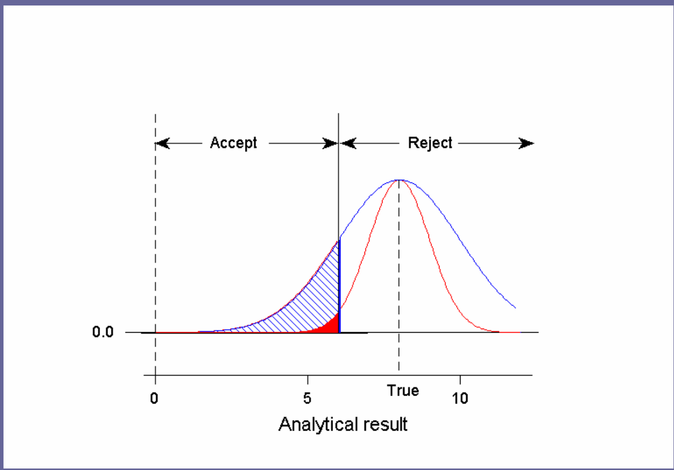
The cost of accuracy, $L_m = f(u)$



Cost of incorrect decisions (L_d): 1—probability of false rejection

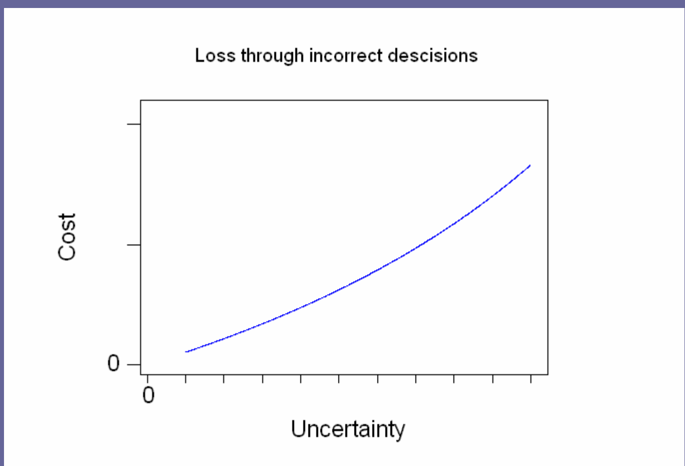


Probabilities of false acceptance

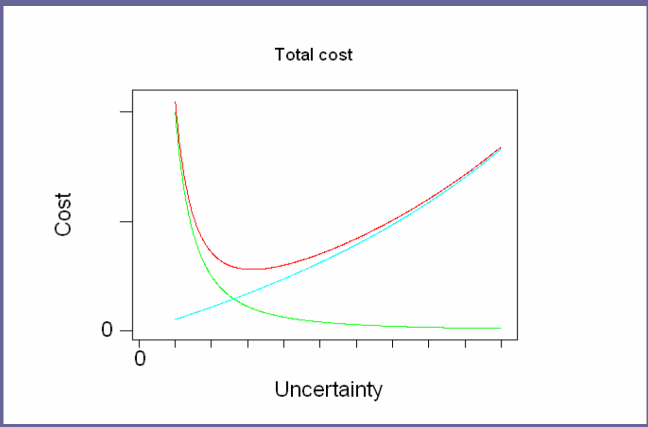


Typical loss function

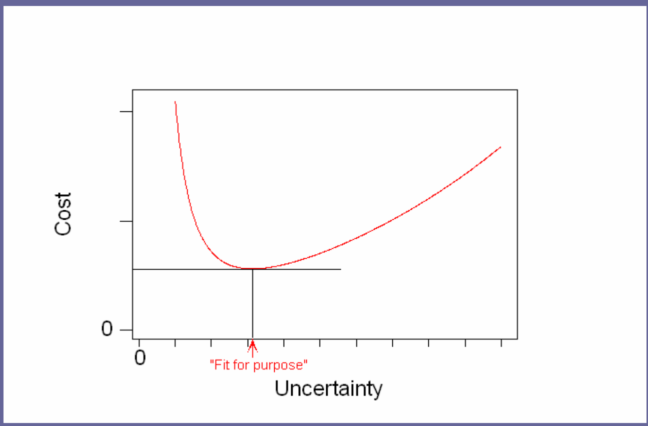
Average loss = Cost of incident × probability of incident



Long-term loss



Fit-for-purpose uncertainty



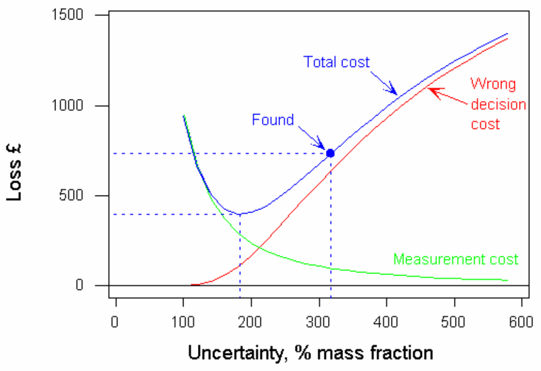
Total cost T

$$T = L_m + L_d$$
$$= L_m + \iint L(x,\mu) P_m(x|\mu) P(\mu)\, \mathrm{d}x\, \mathrm{d}\mu$$

$P_m(x|\mu)$ is the distribution of the result, given the true value;

$P(\mu)$ is the distribution of knowledge about where the true value might lie.

Loss function for sampling lettuce



Getting near-optimal uncertainty

$$\left\{ \begin{aligned} \hat{\sigma}_{anal} &= 168 \\ \hat{\sigma}_{samp} &= 319 \\ \hat{\sigma}_{comb} &= 361 \end{aligned} \right\} \quad \left\{ \begin{aligned} \sigma_{anal} &= 168 \\ \sigma_{samp} &= 160 \\ \sigma_{comb} &= 232 \end{aligned} \right\}$$

Original:
10 increments

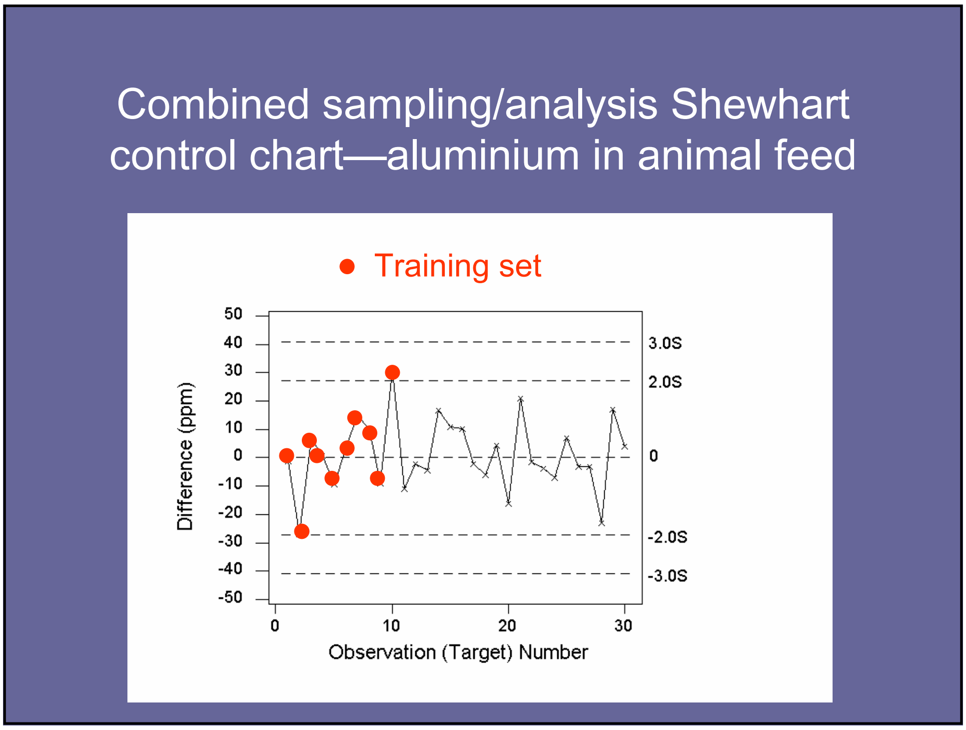
Proposed:
40 increments

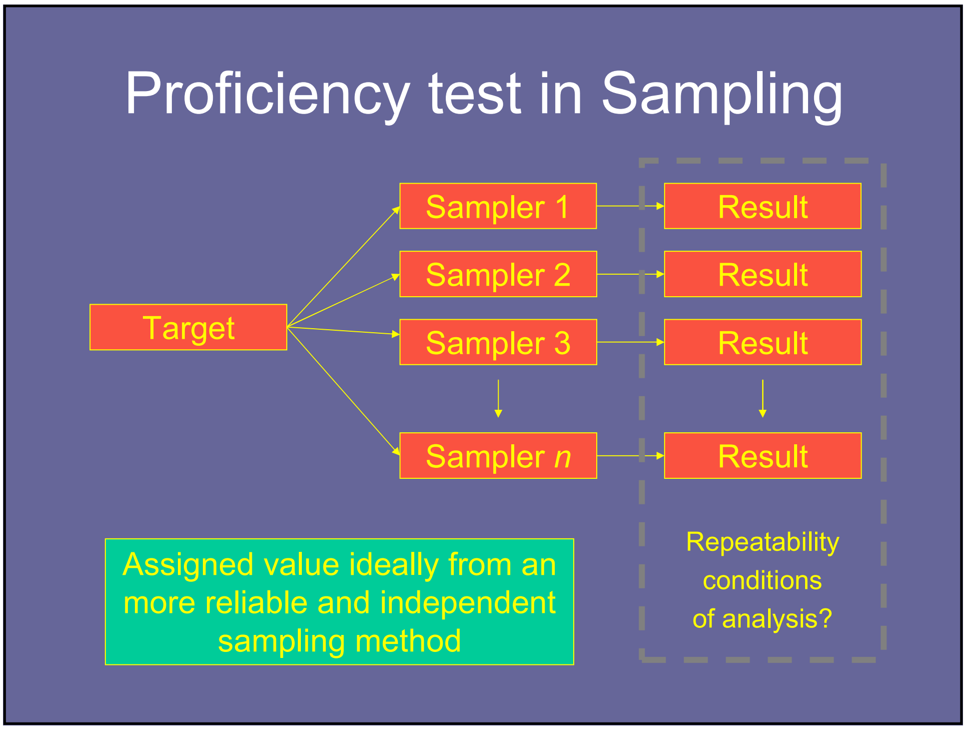
Internal quality control

Target → Sample 1 → Result 1 → Difference d
Target → Sample 2 → Result 1 → Difference d

$$s_d = \sqrt{2(s_{sam}^2 + s_{an}^2)}$$

- Set up control chart for d (Shewhart or J-chart) using control lines at 0, ±2s, and ±3s, where
- For routine or occasional use (for sampling and analysis combined).
- *Note: a result may be unfit for purpose, even if the error is due to heterogeneity and not the method.*


Combined sampling/analysis Shewhart control chart—aluminium in animal feed
Training set
Difference (ppm)
50
40
30
20
10
0
-10
-20
-30
-40
-50
3.0S
2.0S
0
-2.0S
-3.0S
0
10
20
30
Observation (Target) Number


Proficiency test in Sampling
Target
Sampler 1
Sampler 2
Sampler 3
Sampler n
Result
Result
Result
Result
Assigned value ideally from an more reliable and independent sampling method
Repeatability conditions of analysis?

Desiderata

- Samplers use their own preferred sampling protocol.
- Scheme provider conducts analysis under repeatability conditions (with $\sigma_a << \sigma_{s(R)}$).
- Provider specifies a fitness-for-purpose criterion.
- Provider uses an independent assigned value if possible
- Provider calculates a z-score.

Practical points

- Sampling must be "replicable" and "unobserved".
- There may be overall target-specific bias if assigned value is a consensus.
- Expensive.
- Usage: research only at present

General references

- *Measurement uncertainty arising from sampling: a guide to methods and approaches.*
  M H Ramsey and SLR Ellison
  Eurachem/Eurolab/CITAC/Nordtest/ AMC Guide
  2007, 111 pages.
  (Free download from www.eurachem.org/guides/UfS_2007.pdf)

- *Uncertainty from sampling, in the context of fitness for purpose*. (Review)
  M H Ramsey and M Thompson
  Accred Qual Assur, 2007, **12**, 503-513.