

Assessment of the practicality and usefulness of sampling proficiency tests in the food sector

Research programme [Research projects -](#)
Study duration January 2005 to December 2011
Project code E01070

We used real examples (moisture in butter, and patulin in apple juice) and participant feedback on this project.

[Back to top](#)

Background

The quality of measurements made to assess the concentration of components in food is fundamental. The decisions made by producers, vendors, consumers and the regulators of foods rely upon their accuracy. In recent years, it has been increasingly realised that an important part of the measurement process is often the sampling of the food that takes place outside the laboratory.

In a Sampling Proficiency Test (SPT), several organisations ($n \geq 8$) are required to take independent samples, and make analytical measurements, on the same sampling target. This would typically be a bulk consignment of a food material. The analytical measurements can be made by each participant (or their nominee), but the concept of the SPT is tested initially when all the chemical analyses are made in one laboratory.

The proposed SPT on food was the first of its kind in the world, and provided a new tool for assessing the proficiency of samplers. This should help the individual samplers to improve their performance, and become aware of the effect of sample heterogeneity on the measurement result.

Potentially, it should improve the quality of sampling in the whole food sector, which has happened as a consequence of analytical proficiency tests. This should enable better international comparability of measurements, especially if future SPTs are conducted with international participation. Moreover, the results of any SPT will provide information to enable more realistic estimates of the uncertainty of measurements. This, in turn, will improve the reliability of decisions based upon the composition and contamination of foods.

[Back to top](#)

Research Approach

A sampling target was selected to fulfil the necessary criteria, in terms of size, analyte (considering the limit of detection, expected precision, temporal stability, and heterogeneity), availability of defined sampling protocols and common occurrence of commodity.

It was essential to characterise the candidate sampling targets before the SPT in terms of essential characteristics (eg. size, analyte, heterogeneity, limit of detection etc). In order to detect any temporal change, the target will need to be characterised for the target analyte before and after the SPT.

Participants with relevant experience of the sampling target/analyte were selected, given a clear description of their objective and instructions for the SPT. The instruction clearly identified the sampling target and the participants selected whatever protocol they considered most appropriate.

The scoring scheme was devised in general agreement with the harmonised protocol for analytical proficiency tests (Thompson et al. for IUPAC, 2005), but adapted to the particular requirements of a SPT. Participants were supplied with their score with related comments and estimates of the measurement uncertainty.

Comments on both the practicality and usefulness of the SPT from participants and organisers were collated. The constraints on the selection of the sampling target and participants were reviewed and conclusions were reached as to the broad applicability of SPTs in the food sector. The usefulness to the participants were reviewed, and compared with the perceived usefulness of analytical PTs. Lessons learnt from the initial application were addressed in a second application of the SPT in the later stages of the research project.

[Back to top](#)

Results

Two experimental SPTs were held to demonstrate their feasibility, practicality and usefulness in the food sector. The first ever SPT on food looked at the determination of the toxin patulin in cloudy apple juice, using 9 trained participants to sample one batch of 6500 litres of unstirred juice from a tap on the side of the tank. It found that the expanded uncertainty on the measurements (19.4%) was so dominated by the precision of the analytical method, and the juice was so homogeneous, that it was impossible to detect differences between the performances of the samplers. The process of implementing this prototype test was used to identify 12 generic criteria that can be applied to design an effective SPT.

These criteria were then used to design a second improved SPT, on the determination of moisture in fresh butter, which thereby overcame many of the limitations found in the first SPT. The precision of the analytical method was small enough to quantify the heterogeneity of the butter, and the proficiency of the samplers was reflected in their performance z-scores. Using a newly-devised scoring system, two samplers were found to be non-proficient. One of these non-proficient samplers was untrained and intentionally placed in the SPT to test its usefulness, but the other was a trained sampler. The results of the second SPT also demonstrated their usefulness in identifying non-proficient samplers. SPTs therefore have the potential to be used to improve the performance of samplers, in an analogous way to the way in which analytical PTs have already improved the performance of analytical laboratories over recent decades.

A further use of SPT results is to improve estimates of Measurement Uncertainty. It has already become accepted that the sampling procedure contributes to the uncertainty of measurement results in most circumstances, but previous methods of estimation did not include the contribution from sampling bias. The results of the second SPT showed that the MU estimated by the equivalent of the accepted 'duplicate method' (0.39%) was a factor of two lower than the

more realistic value estimated by the SPT (0.87%), which does include substantial between-sampler bias.

The ultimate uptake of SPTs in the food sector may well be driven by their perceived usefulness. These prototype SPTs have been effective in identifying how SPT design needs to improve, and how their usefulness can be assessed. The potentially best scoring scheme, based on the best experimental design, generates 4 z-scores per participant, but combines them in one value (with a technique such as a rescaled sum of z-scores ($RSZ = \sum z/m$)). More research is required to refine this scoring method, so that it is shown to fairly reflect the proficiency of samplers, over a wider range of sampling targets.

[Back to top](#)

Published Papers

Ramsey, M.H., Geelhoed, B., Damant, A.P. and Wood, R. Improved evaluation of measurement uncertainty from sampling by inclusion of between-sampler bias using sampling proficiency testing. *Analyst*, 136, 1313-1321, 2011

Research report

[View Feasibility, Practicality and Usefulness of Sampling Proficiency Tests in the Food Sector as PDF \(4.68 MB\)](#)