On the Collaborative Trial in Sampling

Michael H. Ramsey[a], Ariadni Argyraki[a] and Michael Thompson[b]

[a] *Department of Geology, Imperial College, London, UK SW7 2BP*
[b] *Department of Chemistry, Birkbeck College, University of London, Gordon House, 29 Gordon Square, London, UK WC1H 0PP*

The feasibility of the collaborative trial for validating a sampling protocol is demonstrated. An example in environmental sampling was convenient for the study but the method could be applied in any situation where replicate samples could be taken. A number of samplers each independently take duplicate samples in a random fashion. The samples are then analysed chemically under repeatability conditions and then resultant data analysed statistically by nested analysis of variance. Results can be compared with criteria based on considerations of fitness-for-purpose, and uncertainties associated with sampling can be estimated.

Keywords: Sampling quality; collaborative trial; analysis of variance; fitness-for-purpose; uncertainty

Introduction

This paper is concerned with the validation of sampling protocols by means of an analogue of the collaborative trial (method performance study) used for the validation of analytical protocols. Thousands of analytical methods have been validated by the collaborative trial, for which purpose a definite *modus operandi* has been established.[1,2]

In design the collaborative trial is a replicated experiment in which a number ($n > 4$) of study materials are circulated to the participant laboratories ($m > 7$) for duplicate analysis. The resultant data are treated by analysis of variance (ANOVA) to provide estimates of the repeatability standard deviation ($s_r = s_0$) and the reproducibility standard deviation ($s_R = \sqrt{s_0^2 + s_1^2}$) for each separate study material. The symbols s_0 and s_1 refer to the within-group and the between-group standard deviations quantified by the ANOVA. Repeatability conditions refer to measurements made on a material by one analyst using the same procedure and equipment in one location during a short period of time. Reproducibility conditions refer to measurements made on a material by different analysts using the same procedure in different laboratories.[3] If assigned values for the concentrations of the analyte are also available, the bias of the method for each study material can be estimated, in addition to the precisions.

It is clear that interlaboratory study is necessary for method validation. The interpretation of the method protocol can be quirky or its execution flawed in particular laboratories. Hence the protocol is effectively the manner in which the consensus of laboratories interpret and execute it.

In this study we apply the principle of the collaborative trial to sampling methodology and demonstrate the feasibility of the collaborative trial in sampling (CTS). Although the example chosen for study relates to environmental science, the CTS would be applicable to any situation where sampling can be replicated (which might exclude some flowing materials).

The technical requirements and applicability of CTS have been discussed previously in the context of the whole field of quality in sampling.[4] The approach considered here is similar to the analytical collaborative trial. A number of samplers independently take duplicate samples at random from the sampling target (*i.e.*, the object, lot of material or site to be sampled). The samples are then analysed in duplicate under randomized repeatability conditions (*i.e.*, in a single run in one laboratory). The resultant data are analysed statistically by nested ANOVA to provide estimates of the analytical repeatability standard deviation ($s_{r(a)} = s_0$), the sampling repeatability standard deviation ($s_{r(s)} = s_1$) and the between-sampler standard deviation (s_2). The sampling reproducibility standard deviation is estimated as $s_{R(s)} = \sqrt{s_1^2 + s_2^2}$. Hence $s_{r(s)}$ quantifies the variation between the samples collected from a single target by one sampler using the same protocol and equipment within a short period of time, while $s_{R(s)}$ quantifies the variation between samples collected from a single target by different samplers using different equipment but according to the same protocol. Chemical analysis under repeatability conditions is necessary to avoid confounding analytical variations with sampling variations. The term 'sampling' is defined here to cover all of the processes up to and including the preparation of the laboratory sample. The process of 'analysis' begins with the sub-sampling of the laboratory sample, *i.e.*, the weighing out of the test portion.

Experimental

Sampling Target

Only one target was used in this study. It was a sub-rectangular field of about 60 × 150 m near Wirksworth in Derbyshire and had been subjected to pollution from a small lead smelter between the 14th and the 16th Centuries. The field had not been ploughed for at least 45 years and was currently in use for grazing horses. Previous study had shown that the field was unevenly contaminated with lead and other elements associated with the smelter, but otherwise fairly uniform.[5]

Samplers

Nine organizations listed in the acknowledgement (two commercial organizations, one government laboratory and six university departments) sent samplers to the site over the course of one week in August 1994. The samplers were given an explanation of the aims and intentions of the project one month before the sampling period, and the exact protocol immediately before commencing sampling. No sampler observed any other during the sampling exercise.

Sampling Protocol

Equipment

The equipment used comprised the following: soil auger, steel, 2.5 cm diameter; sample bags, wet strength, trace-element free; marker canes (optional); surveying tape (optional); indelible marker pen.

Method

The method studied is frequently used by soil samplers,[6–8] and was employed here as an example that could be quickly and simply executed rather than because of any special performance capability. The term 'increment' is recognized to mean the portions of the sampling target, collected at various points, that are combined to make up the 'aggregate sample'. The part of the aggregate sample that is, after mechanical treatment such as comminution and dividing, used for analysis is called the laboratory sample.[9]

(*i*) Sample the field using a design in the shape of a 'W' (see Fig. 1). Select the corner to commence sampling and the orientation of the 'W' by random numbers.

(*ii*) Collect the increments at the vertices of the 'W' and at three roughly equidistant points along each leg, giving 17 increments in total. Use canes as position markers if preferred.

(*iii*) Use the auger to take the increments of soil to a depth of 15 cm. To do this first remove the surface vegetation from the sampling point and, holding the auger vertically, screw the blade into the soil to a depth of 15 cm. Pull out the auger without rotation, trying to avoid smearing. Remove the soil from the auger blade and place it in the sample bag. Cover the auger hole, as far as possible leaving no trace that material has been collected.

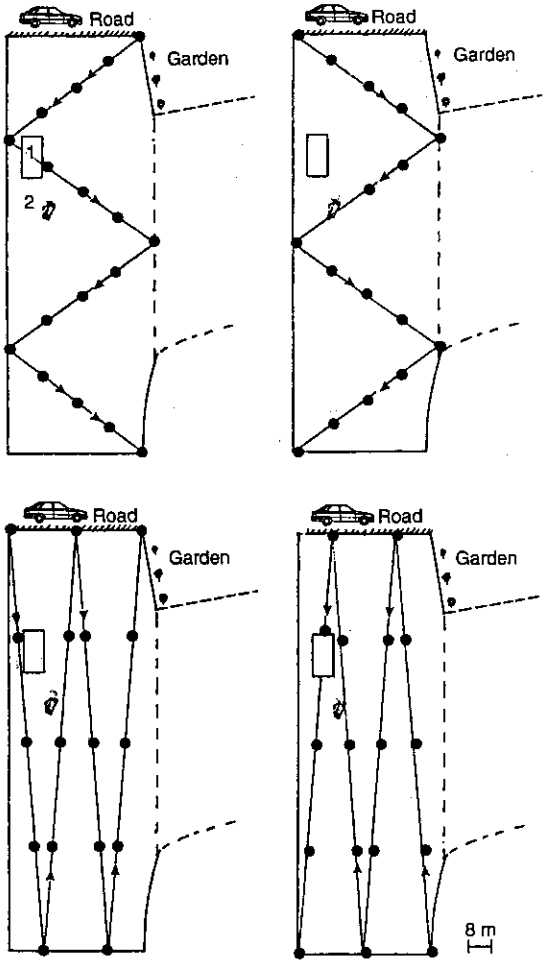

Fig. 1 Map of the sampling target, showing the four possible 'W' patterns for collecting the increments. 1, Old shed; 2, abandoned agricultural machine; filled circle, sampling point.

(*iv*) Follow the same procedure for each increment, combining the increments in the one bag to make the aggregate sample.

(*v*) Collect a duplicate sample by repeating the whole procedure [steps (*i*)–(*iv*)]. Select a new starting corner and orientation by using a second set of random numbers. Collect the duplicate aggregate sample in a separate bag.

Mechanical Preparation of the Samples

The aggregate samples were collected by the organizers for mechanical preparation and chemical analysis. Each aggregate sample was dried at 65 °C and broken down to individual grains with a pestle and mortar. The fraction passing a 2 mm plastic sieve was ground in a Tema mill to <100 μm to comprise the laboratory sample.

Chemical Analysis

The chemical analysis was carried out on duplicate test portions of each laboratory sample, in a random order under repeatability conditions, *i.e.*, within a single run. (Duplicate analyses were not made consecutively but at a random position within the run, so as to give an unbiased estimate of the repeatability precision.) The test portions (0.1 g) were treated with a mixture of nitric and perchloric acids to solubilize the analytes which were presented for analysis in dilute hydrochloric acid solution.[10] Copper and lead were determined in the solutions by ICP-AES. Internal quality control was provided by the inclusion in the run of six appropriate reference materials.

Statistical Analysis

The standard deviations s_0, s_1, s_2 were estimated by classical ANOVA and by the robust ANOVA method recommended by the Analytical Methods Committee (AMC)[11] using a program adapted for the nested design. The robust method reduces the influence of outlying results which would otherwise have a disproportionate influence on the standard deviations. The elimination of the influence of outliers is widely regarded as essential in analytical collaborative trials. The AMC robust method gives results similar to the 'harmonized protocol' method[2] (the rejection of outliers followed by classical ANOVA) but is rather quicker to execute and avoids the need to identify outliers.

Results and Discussion

The design of the experiment is shown in Fig. 2: each sampler collected random duplicate samples according to the sampling protocol, and the samples were analysed under randomized repeatability conditions. The elemental concentrations determined in the laboratory samples are recorded in Table 1 and presented in Fig. 3. The statistics are given in Table 2. The order in which the data are presented differs from that of the list of participants below. The analytical quality control data are represented in Table 3.

A small analytical bias was evident in the results for both lead and copper, but that would not affect the outcome of the sampling trial. A relatively large percentage bias (40–60%) was detected at low concentrations of lead (30–40 μg g^{-1}). That is equivalent to an absolute bias of about 12 μg g^{-1} and is probably caused by instrumental memory effects from the very high concentrations in the samples. In the concentration range of the samples (4000–10 000 μg g^{-1}) the relative bias detected for lead was less than 5%. Moreover, the fact that all of the samples were analysed at random within one run makes any bias detected irrelevant to the objectives of this study.

Differences in the classical and the robust statistics are apparent but, in this instance, are not clearly of any moment. However, it must be remembered that some sampling targets are intrinsically heterogeneous. Outlying samples would be a normal expectation in these circumstances and the precisions estimated should refect this. Hence there may be compelling arguments in some CTS to include outlying results in the calculation of the standard deviations, so long as they reflect differences between samples rather than differences between analyses. Such decisions require detailed consideration of the data themselves and the purpose to which the data will be put. A degree of statistical expertise is required for the task. It would be unwise to generalize on the treatment of outliers until considerably more experience with CTS has been accumulated.

The data for copper show that the element is evenly distributed throughout the plot. As a consequence the sampling standard deviations are small and comparable with the analytical error. The sampling reproducibility is noticeably higher than the sampling repeatability, *i.e.*, there are noticeably greater differences between the samplers than is explained by variations in the performance of a single sampler. This is similar to the analogous situation in chemical analysis, where interlaboratory variations tend to be greater than those within a laboratory. The sampling relative standard deviations are reasonably small (for the particular concentration of copper), and for most environmental or agricultural studies the sampling protocol could be regarded as fit for purpose. The sampling statistics could be used if required to provide an estimate of the uncertainty of sampling at a single concentration of the analyte.

In contrast to the foregoing, the results for lead reflect the extreme variations in its concentration over the plot. Although the analysis is uniformly of good precision (as can be judged from the analytical duplicate results and the analytical relative standard deviation of 3.7%) differences between duplicate samples are large giving rise to a sampling repeata-

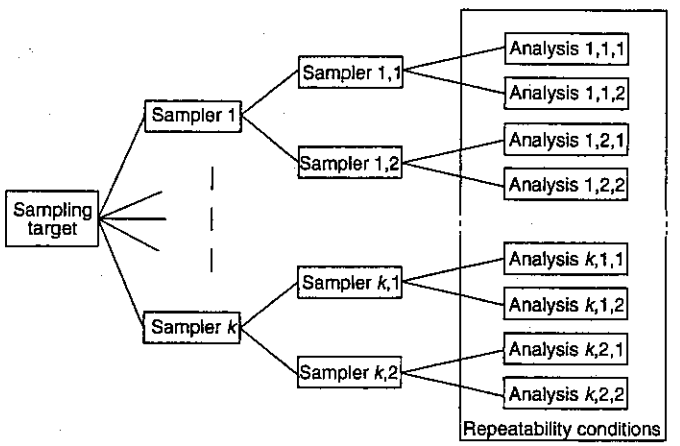

Fig. 2 Schematic design of the collaborative trial in sampling. For further details see Fig. 6, of ref. 4.

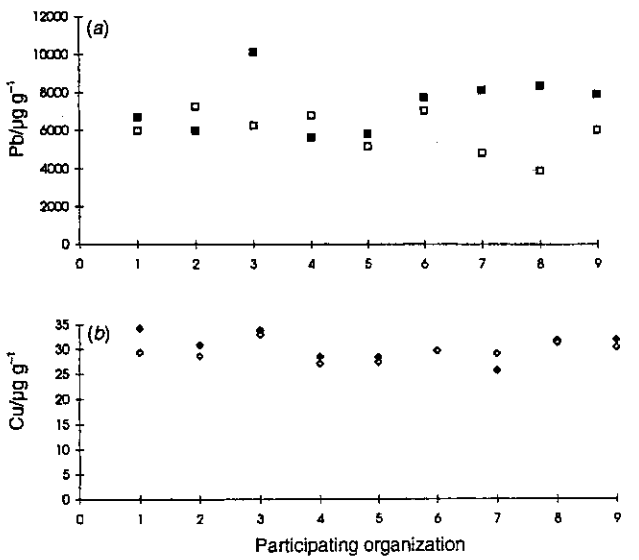

Fig. 3 Graphic representation of the results of the collaborative trial in sampling. Each point shows the mean of the two analytical duplicates.

Table 1 Data from the sampling collaborative trial (μg g^{-1})

Sampler	Sample	Analysis	Cu	Pb
1	1	1	35.3	5871
		2	33.0	6114
	2	1	30.0	6842
		2	28.6	6541
2	1	1	31.3	7366
		2	30.2	7152
	2	1	28.4	6079
		2	28.8	5890
3	1	1	32.8	6113
		2	34.9	6369
	2	1	32.8	9826
		2	32.9	10451
4	1	1	28.2	6707
		2	28.7	6867
	2	1	26.4	5647
		2	27.8	5582
5	1	1	28.9	5218
		2	27.7	5102
	2	1	27.7	5953
		2	26.9	5642
6	1	1	29.9	7016
		2	29.3	7015
	2	1	30.6	7978
		2	28.8	7455
7	1	1	26.1	4779
		2	25.1	4869
	2	1	28.6	8166
		2	29.4	8015
8	1	1	32.1	3771
		2	31.1	3964
	2	1	32.9	8271
		2	29.5	8359
9	1	1	32.5	6043
		2	31.0	5977
	2	1	31.3	7915
		2	29.3	7851

Table 2 Statistics from the sampling collaborative trial (μg g^{-1})

	Cu		Pb	
Statistic	Classical	Robust	Classical	Robust
Grand mean	30.0	30.0	6618	6575
s_0	1.1	1.1	245	165
$s_1 = s_{r(s)}$	1.4	1.2	1683	1822
s_2	1.7	2.1	0	0
$s_{R(s)}$	2.2	2.4	1683	1822
RSD_0	3.7%	3.7%	3.7%	2.5%
$RSD_{r(s)}$	4.7%	4.0%	25.4%	27.7%
$RSD_{R(s)}$	7.3%	8.0%	25.4%	27.7%

Table 3 Results of the analysis of reference materials (μg g^{-1})

CRM	BCR141	BCR142	BCR143	NBS2709	NBS2710	NBS2711
Pb—						
n	3	3	3	3	3	3
x	42.22	43.66	1312.59	31.05	5339.55	1179.62
s	11.75	4	0.95	15.02	0.5	2.89
Cert. val.*	29.4	37.8	1333	18.9	5532	1162
Bias	12.82	5.86	−20.41	12.15	−192.45	17.6
Percentage bias	43.6	16	−2	64	−4	2
Cu—						
n	3	3	3	3	3	3
x	30.6	25.64	227.8	33.38	2776.88	108.68
s	6.41	3.8	0.28	1.58	0.96	1.95
Cert. val.*	32.6	27.5	236.5	34.6	2950	114
Bias	−2	−1.9	−8.7	−1.2	−173.1	−5.32
Percentage bias	−6	−7	−4	−4	−6	−5

Cert. val. = certified value.

bility relative standard deviation of 25.4%. The fitness-for-purpose of the sampling protocol might therefore be questionable. Variations between duplicate samples are so great that in this instance no significant difference between samplers could be detected. There are no obvious grounds for using robust statistics in this example.

Conclusions

Despite the limited scope of the experiment described, the feasibility and usefulness of the CTS has been demonstrated. In a fully developed CTS (*i.e.*, with several different sampling targets perhaps with varying average concentrations of analyte) it is clear that the performance of a sampling protocol could be evaluated and compared with criteria based on fitness-for-purpose considerations. Estimates of the uncertainty associated with the sampling protocol could be made. If assigned values could be attributed to the sampling targets (which may be possible in some instances[7]) then estimates of the sampling bias could also be obtained.

By considering two analytes with very different distributions within the sampling target, we have shown that between-sampler variations are sometimes apparent, but may be obscured under some conditions (*e.g.*, with a grossly heterogeneous target). The comparable situation is unusual in analytical collaborative trials, and this highlights the lack of complete homology between such trials in sampling and in chemical analysis. The contrast stems in part from the absence in sampling of calibration, one of the main contributors to interlaboratory (analytical) variations. The more important contribution, however, is probably the heterogeneity of the sampling target, which will be an essential and often conspicuous feature of the CTS. In an analytical trial the organizer is careful to homogenize the trial materials, so that the contribution of heterogeneity to the reproducibility standard deviation is usually negligible.

For lead the demonstrated large sampling precisions in comparison with the analytical precision must be seen as reinforcing a strong warning that, in many instances, the uncertainty of sampling may dominate the total uncertainty of measurement and render analytical excellence futile. The contrast between the results for copper and lead also emphasizes a fact that is obvious in principle, but which may still be overlooked: a sampling protocol applied to a particular type of target could be fit for purpose for one analyte but not for another.

In the study described, the protocol has been applied at only one level, *i.e.*, at only one concentration of each analyte, and on only one example of the target material. Even if the analyte concentration were normally within a narrow range, it would still be good practice to repeat the trial on several distinct targets to ensure that the conclusions drawn were reasonably general. A single target might turn out to be atypical in some unforeseen way. Where the concentration of the analyte is likely to vary over a large range between targets, the validation of the protocol at different concentrations would be essential, so that changes in sampling repeatability and reproducibility with concentration could be documented.

While the feasibility and potential usefulness of the CTS have been demonstrated, considerably more experience with the practical aspects is needed before any guidance can be provided on optimum methodology for a trial in any particular sector of activity. The authors would be interested to hear from and collaborate with readers who would like to pursue questions of sampling quality related to this paper.

The authors thank the following people and organizations for participating in this study: P. Nathanail of Wimpey Environmental; J. Nathanail of Delta Environmental; M. J. Brown of the British Geological Survey; M. Hill of the Department of Environmental Sciences, University of Bradford; J. Rae of the Postgraduate Institute for Sedimentology, University of Reading: P. Abrahams of the Department of Geography, University College of Wales: J. Cotter-Howells of the Department of Geology, University of Manchester; and A. Mehra of the School of Environmental and Applied Sciences, University of Derby.

References

1 *Precision of Test Methods* (currently under revision), ISO 5725: 1986, ISO, Geneva, 1986.
2 Horwitz, W., *Pure Appl. Chem.*, 1988, **60**, 855.
3 *International Vocabulary for Basic and General Terms in Metrology (VIM)*, ISO, Geneva, 2nd edn., 1993.
4 Thompson, M., and Ramsey, M. H., *Analyst*, 1995, **120**, 261.
5 Maskall, J., and Thornton, I., *Land Contamination and Reclamation*, 1993, **1**, 92.
6 *Sampling Soils for Analysis*, Leaflet 655, Ministry of Agriculture, Fisheries and Food, London, 1979.
7 Ramsey, M. H., Argyraki, A., and Thompson, M., *Analyst*, 1995, **120**, 1353.
8 Thompson, M., and Maguire, M., *Analyst*, 1993, **118**, 1107.
9 Horwitz, W., *Pure Appl. Chem.*, 1990, **62**, 1193.
10 Thompson, M., and Walsh, J. N., *A Handbook of Inductively Coupled Plasma Spectrometry*, Blackie, Glasgow, 1989, pp. 143–144.
11 Analytical Methods Committee, *Analyst*, 1989, **114**, 1693.

Paper 5/00756I
Received February 8, 1995
Accepted June 5, 1995