ILAC Discussion Paper on Homogeneity and Stability Testing

ILAC Proficiency Testing Consultative Group, Task Force on H&S testing:
Dan Tholen, Convenor, A2LA
Nick Boley, LGC
Jane Gun-Monro QMP-LS
Jeff Gust, Quametec PT
Wolfgang Korth, National Residue Survey
Michael Noble, CMPT
Penny Petinos, RCPA
Dan Wruck, QHHS

I. Scope

This discussion document is intended to present the technical issues concerning the homogeneity and stability of proficiency test samples. It is intended also to lead to consensus recommendations concerning the resolution of these issues. These recommendations have served as guidance for the revisions of ILAC G13 and ISO/IEC Guide 43 (ISO/IEC 17043) and for use by ILAC member bodies in assessing the suitability of PT schemes used by accredited laboratories.

II. Introduction

International guidelines for the accreditation of providers of proficiency testing or external quality assessment schemes (PT) require providers to demonstrate both the homogeneity and the stability of proficiency testing samples. This is essential if PT is to serve its purpose and be able to identify errors in testing or calibration and not have false warnings due to inhomogeneous or unstable samples.

There are many parameters to consider when assuring quality PT samples. Obviously, one has greater confidence in the samples if the provider has significant experience with the particular program, and has documented records of homogeneous and stable samples. Accreditation of the PT provider or certification of the manufacturer of samples also adds confidence in the ability to provide suitable samples.

Nevertheless, every accredited PT scheme provider must be able to justify the claim that its samples are homogenous and stable. Requirements vary greatly for different accrediting bodies and in different fields of testing and calibration. In some situations, homogeneity and stability testing is required for every analyte in every lot of samples. Providers of other schemes may only test indicator analytes, or may conduct testing at regular intervals but not for every lot of proficiency testing samples.

In some areas such as non-destructive testing, field testing, or testing calibration artifacts, the ‘sample’ may be identical for all participants. In such cases the PT scheme provider has procedures to assure that the sample is stable throughout all test runs and an appropriate homogeneity is assured for all measurements within a defined range. Artifacts that are transported between participant laboratories need to be checked using defined procedures to ensure that the assigned values for the artifact have not changed.

It is usual to assess both homogeneity and stability by manufacturing and testing samples additional to those required for distribution to participants. Regardless of the extent of homogeneity and stability testing, it can be expensive, and it uses samples that could be used for other purposes. Expense can be of critical importance if either the sample or the test method is expensive or time intensive, especially if the number of participants is small. The issue is further complicated when the PT provider itself does not have the analytical capability to test the samples and where the laboratories that do have the capability are participants in the PT study. The supply of samples can be critical where access to them is difficult. There may be situations where the risk of inhomogeneity or instability is acceptably low, and the potential cost associated with this risk is less than the actual cost of performing the necessary homogeneity and stability testing.

It is also important to consider the effectiveness of homogeneity and stability testing. Usually only a small proportion of samples from an entire lot of production are tested. It is therefore possible that the tests will fail to detect problems that exist in a portion of the total number of samples. Thus a proficiency testing sample may actually be unacceptable, even though it has not been demonstrated by the recommended homogeneity and stability testing.

III. Discussion of Homogeneity and Stability testing procedures

Three situations are discussed:
A. Samples are prepared in bulk, placed into aliquots and distributed to participants.
B. Samples are prepared separately for each participant, all to the same design.
C. The same test item is sent to all participants (this includes audits).

This is followed by several discussions of **general techniques for data analysis:**

1. Accommodating known inhomogeneity.
2. Accommodating known instability
3. Statistical tests for inhomogeneity.
4. Statistical tests for instability.

It is recognized that medical proficiency testing utilizes a great variety of materials, some of which are complex and may be rare. There also are situations where samples must receive special handling so they cannot be treated as normal client samples. However the demands of medical PT are rarely, if ever, confined to the medical sector. The general challenges of scarcity, instability, assumed homogeneity (in some cases), unavoidable heterogeneity (in some cases), expense of testing, artificial samples, etc. are problems shared by other sectors as well. Where applicable, any unique problems that occur in just one field of testing are highlighted.

It is also recognized that some types of samples do not require homogeneity or stability testing (for example, reproduced images). On a case-by-case basis, there may be samples or analytes types that do not require homogeneity or stability testing for justifiable reason. However, many situations of qualitative testing, or presence/absence testing, do require homogeneity and stability testing of some kind.

A. Bulk Samples

Issues to consider for the bulk design:

1. Conventional designs for testing homogeneity and stability.
2. Variations for different numbers of participants.
3. Test every analyte, or indicator analytes.
4. Test every lot (sample) or periodic lots.
5. Types of samples where homogeneity can be assumed.
6. Types of samples where stability can be assumed.
7. Special considerations for microbiology.

A.1. Conventional procedures

A.1.1. Standard method of testing homogeneity. These are well described currently in ISO 13528 (2005), in the IUPAC Harmonized Protocol (2006) and in ISO Guide 35 (2006). In general, the preferred technique is to use duplicate measurements on a minimum of 10 randomly chosen samples. The between sample variability is determined and compared with either a fraction of the evaluation criterion (13528), a statistical test (Guide 35), or a combination of fitness for purpose and statistical criteria (IUPAC).

A.1.2. Standard method of testing stability. Procedures for testing stability are not as well discussed as are procedures for testing homogeneity. However the general procedure is to take a small number of samples at the end of the test period, take a measurement on each one, and compare the mean result with the average of the results determined in the homogeneity check. If the measurement procedure is stable over that time period, this provides the most sensitive indicator of changes in the test item. Conventional statistical tests include looking at a fraction of the evaluation criterion (13528) or a statistical confidence interval placed around the homogeneity mean.

A.2. Considerations for numbers of samples for testing homogeneity

When 100 or more samples are prepared, it is generally agreed that at least 10 samples should be tested for homogeneity. All quoted references (Guide 35, 13528, IUPAC) call for a minimum of 10 samples. Some procedures say to test more than 10 samples, such as the $3*\sqrt[3]{N}$ (with N= total number of samples produced). This would leave the following requirements:

Samples	Tested
1-38	10
39-49	11
50-65	12
100	14
400	22
1000	30

When there are fewer samples, should 10 still be tested? The finite population correction allows us to reduce a variance by 1-(n/N) when n (the number tested) is large relative to N (the number samples made). This could allow homogeneity testing on a smaller number of samples without loss of power to detect inhomogeneity. Alternatively, the choice of number of samples could be left to the provider, who would need to be able to justify the choice. One example reduced scheme:

Samples	Tested
1-2	1
3-9	2
10-19	3
20-39	4
40-59	5
60-79	7
80-99	8
100+	10

A.3. Test Indicator Analytes

Can criteria be developed to affirm that one analyte can stand as an indicator of the homogeneity of other analytes, or of the entire lot of samples? There should be technical knowledge and experience that a group of analytes in a sample will behave (e.g., mix) similarly. This experience should be backed with sufficient evidence. If one indicator analyte is used to stand for many, then if possible the indicator analyte should rotate among the possible analytes.

For example, with a liquid, homogeneity testing on one key measurand would indicate homogeneity for all measurands, with the exception of, for example, dissolved gases, or suspended solids, which need treating individually. For solid materials, it is frequently not feasible to test the homogeneity on all measurands in the test sample, mainly for reasons of cost and time. Indicator measurands must be used. The criteria for these can be based on sound scientific principles (e.g. measurand *a* can be linked to measurands *b* and *c*), and this needs to be discussed and agreed by the technical advisory or steering group. Other criteria, where total homogeneity testing is not feasible, can be based on the importance of the measurand to the population of participants; this may be technical (i.e. the measurand is important for trade or technical reasons) or statistical (i.e. if very few participants routinely test for a particular measurand homogeneity testing may be unnecessary).

Other candidate analytes would be those most sensitive to sample treatment or processing (e.g., mixing), or analytes that are present at very low concentration.

A.4. Test every lot or sample lots

Some providers have experience with a manufacturing process and have produced many PT samples over a long period of time. Data confirm that the process is stable and produces homogeneous samples. Sometimes the management system for these processes is certified to ISO 9001. Can homogeneity be assumed for lots manufactured with identical controlled processes? Some in the group expressed concern about this assumption, and assert that all lots should be tested (as does ISO Guide 35), whilst others believed the process to be valid, particularly if supported by direct homogeneity testing on a validated procedure periodically, covering all analytes over time.

A.5. Homogeneous by design

Can criteria be established to define analytes and matrices where homogeneity can be assumed (e.g., chloride in water, hemoglobin in blood)? Can this just be by technical consensus? For example, simple liquids (e.g., prepared aqueous solutions of metals or other inorganic species) are intrinsically homogenous. For more complex liquids (e.g., drinks) which may be supplied in individual commercial units, or are prepared by bulking many units, homogeneity testing must be carried out to ensure the consistency within a batch, or the efficacy of the mixing process. This can frequently be done with a limited set of measurands.

Caution should be applied here too as some metals may precipitate from solution if pH changes occur during transport, and again the decision about whether the solutions are homogeneous by design should be made by technical experts advising on the particular study. Where they are homogeneous by nature (or design), the limiting factor is the precision of the test method used to demonstrate homogeneity.

Annex 1 presents several considerations for these “true solutions.” This annex includes a discussion on the use of participant results to determine sample suitability.

A.6. Stable by design

Can criteria be established to define analytes and matrices where stability can be assumed (e.g., chloride in water)? Can this be determined by technical consensus? How much supporting data are needed? Supporting data on samples “known” to be intrinsically stable should be available, usually from the literature, or studies previously undertaken by the PT provider, its collaborators or other experienced organisations in the field, including PT participants.

For PT schemes operated on a regular basis with similar sample matrices, it is recommended that stability studies be carried out on a typical sample matrix before commencing the scheme (where data such as that described above are not readily available). If the matrix is shown to be stable for *at least* the normal period of the proficiency test (i.e., from the date of production to the final date for measuring the results), then it can be safely assumed that all similar samples used in the PT scheme from that time on are stable, although occasional validation of this may be necessary. Where possible, however, stability of samples should be known for a longer period, if the PT provider makes surplus samples available as quality control (QC) materials, or for PT participants to re-test following unsatisfactory performance.

Annex 1 has a discussion paper on this topic, advocating three situations where stability testing is not necessary.

A.7. Special considerations for microbiology

Homogeneity and stability testing are appropriate only for tests and measurement involving detectable agents that may be unevenly distributed in samples. Electronic and paper samples can be controlled for quality. It is difficult to determine the homogeneity of conventional control products, but viability should be confirmed. Stability must be checked under actual stress conditions. That is, the stability test should be on samples that were subjected to stresses of shipment and lack of refrigeration. Homogeneity and stability criteria must take into account the possibly greater variability in distribution, and the possibility of containing analytes (or organisms) that were not expected.

B. Considerations for Custom PT samples

B.1 Many PT studies involve a small number of laboratories (generally < 10) and although it is possible for the PT provider to prepare appropriate samples, it may be difficult for the PT provider to test the samples if the provider does not have appropriate analytical facilities. Furthermore, with small numbers, the only labs that have the required analytical capability may also be participants of the study, consequently sending one of these laboratories the PT samples for direct homogeneity testing and the same samples again as part of the PT study might give that lab an unfair advantage over the other participants in the study. Also, in studies involving a small number of participants, the number of samples that would have to be analysed in order to demonstrate homogeneity as is currently required in the "traditional" approach would greatly exceed the number required for the PT participants. The cost of directly testing all PT study samples for homogeneity in this manner quickly becomes prohibitive.

B.2 Two approaches are considered to demonstrate sufficient homogeneity for samples prepared from bulk materials. These approaches are summarised below (Ref. 7 & 8):

Approach (1): One can demonstrate sufficient homogeneity by preparing an additional set of samples at the same time and under identical conditions to the PT study samples. The homogeneity samples involve all analytes specified for the study and each analyte was present at approximately the same concentration. A subset of samples (n=10) is supplied to one of the participating labs for homogeneity testing. Because the results of the homogeneity tests (run according to internationally accepted practice as outlined in Fearn and Thompson, 2001, Analyst 126:1414-1417) showed that they were sufficiently homogeneous, it was assumed that the PT study samples prepared at the same time and under identical conditions, but containing different concentrations of the target analytes, were also homogeneous. These PT study samples were subsequently sent to all laboratories, including the lab that analysed the prior homogeneity test samples, for analysis.

Approach (2): There is also a process where samples are prepared according to the relevant standard operating procedure (SOP) for the preparation of PT samples for a particular study (e.g., pesticides in fat) and these samples are sent to a lab for homogeneity testing. On the basis that the results of the homogeneity tests demonstrate that the sample preparation procedure is capable of generating samples of sufficient "fit-for-purpose" homogeneity, subsequent samples prepared using the same technique are assumed to be homogeneous unless there is direct evidence otherwise (from the PT results).

B.3 The discussion paper in Annex 2 asserts that custom-made samples should not be tested for homogeneity.

C. All participants see the same sample (and audit samples)

C.1. For this section, a paper was provided by Quametec PT that discusses the relevant issues for samples that are sent out sequentially by the coordinator or consecutively from participant to participant. This discussion paper is given as an annex.

C.2. The paper relates to many situations in which a single sample is used to audit a laboratory's competence. This might be called "audit" samples or "one-off' tests. These are situations where a sample has an accepted reference value and an evaluation range that are assumed to be correct. Therefore the concerns for storage, shipping, and appropriate re-check of the stability all apply.

IV. General Statistical Concerns

1. Accommodating known inhomogeneity
If samples are known to be inhomogeneous, but are considered useful for the program, the results can still be used. The heterogeneity in the samples is observed as increased differences between laboratories. If the between laboratory variability is used in the evaluation (as in conventional z scores) then the scores already account for inhomogeneity. If a fitness for purpose criterion is used, then the inhomogeneity variance should be added to the evaluation criterion in an appropriate mathematical way, depending on the nature of the evaluation limit. Any estimate of the uncertainty of the assigned value must take this inhomogeneity into account.

If there is known inhomogeneity in a sample in some analytes, but not all (say for one or two measurands out of ten tested), the sample can still be distributed, providing the inhomogeneity is not very significant (to be defined). In such cases, the statistical protocol should states that such inhomogeneity is taken into account when calculating z-scores. It may occur that the "known" inhomogeneity is not verified by the participants' results.

2. Accommodating known instability

If samples are known to be unstable and the instability cannot be controlled, then they should not be used in a PT study. The instability must be controlled either by limiting the time and conditions of analysis for every laboratory, or by using a change function to predict change at the time of analysis.

3. Statistical tests for inhomogeneity

A. F test (10+ samples) (Guide 35)

B. Standard Deviation (SD) sample / σ_{PT} ratio (ISO 13528)

C. SD sample expanded with F test (new IUPAC)

If a sample cannot be tested in duplicate, homogeneity is checked with option B. Option C may also be used if the laboratory has an independent estimate of repeatability.

4. Statistical tests for instability

A. Statistical t test of stability vs. homogeneity results

B. Change based on size relative to σ_{PT}

Annex 1: Guidelines and Requirements

<u>ISO Guide 43-1</u>:
5.6.2 *Where bulk material is prepared for a proficiency test, it should be sufficiently homogeneous for each test parameter so that all laboratories will receive test items that do not differ significantly in the parameters to be measured. The coordinator should document the procedure used to establish the homogeneity of the test item.*

When possible, homogeneity testing should be carried out prior to dispatch of the test items to the participating laboratories. The degree of homogeneity should be such that differences between test items will not significantly affect the evaluation of a participant's result.

5.6.3 *Where possible, the coordinator should also provide evidence that the test items are sufficiently stable to ensure that they will not undergo any significant change throughout the conduct of the proficiency test. When unstable measurands need to be assessed, it may be necessary for the coordinating organization to specify a date by which the testing should be completed, and any special pre-testing procedures.*

A.4 *Appropriate statistical techniques should be used for the evaluation of data from homogeneity testing of test items.*

<u>ILAC G13:2007</u>

3.3.2.3 The provider shall be able to demonstrate that the proficiency test items are sufficiently homogeneous for the particular proficiency testing scheme.

Note 1: A relatively inhomogeneous material may be the best available, and may therefore still be useful as a proficiency test material provided the uncertainties of the assigned property values take due account of this.

Note 2: In some cases it is not feasible for proficiency test items to be subjected to homogeneity and stability testing. Such cases would include, for example, where proficiency test items are taken from patients, with limited amounts of material available. In these circumstances the proficiency test provider should have evidence to demonstrate that the procedures used to collect, package and distribute the proficiency test items are capable of maintaining homogeneity and stability, or some other form of justification.

3.3.3 Homogeneity and stability testing

3.3.3.1 Where appropriate, the provider or its subcontractors shall use a statistically random selection of a representative number of samples from a batch of test material to assess the homogeneity of the material.

This assessment procedure shall be documented and be conducted, where applicable, in accordance with acceptable statistical designs.

3.3.3.2 The assessment of homogeneity should be performed after the proficiency test items have been packaged in the final form and before distribution to participants unless, for example, stability studies indicate that it should be stored in bulk form. In some cases, an intermediate homogeneity check may be necessary, for example, before sealing into ampoules.

Note: Homogeneity testing may on some occasions not be done prior to distribution for practical, technical, or logistical reasons, but great caution must be exercised if it is not done or if it is done after test results have been received. In all cases, the provider is required to document the procedure by which it is ensured that homogeneity is adequate.

3.3.3.3 Where appropriate, the property values to be determined in the proficiency testing scheme shall be measured periodically, preferably over a range of conditions under which the proficiency test item is to be stored prior to distribution.

3.3.3.4 Proficiency test items shall be demonstrated to be sufficiently stable to ensure that they will not undergo any significant change throughout the conduct of the proficiency test, including storage and transport conditions.

Note 1: If the proficiency test item is used for schemes extending over a lengthy period of time, then, depending on the nature of the item, it may also be necessary to carry out stability checks during the period of use.

Note 2: Criteria for suitable homogeneity and stability should be based on the effect that heterogeneity and instability will have on the uncertainty of the participant's result, and thereby on the evaluation of the acceptability of a participant's result.

3.3.4.2 Appropriate statistical design of a proficiency testing scheme is essential. In designing a scheme the provider shall give careful consideration to the following:

h) where appropriate, the homogeneity and stability of proficiency test items.

3.6.1.7 The provider shall have documented criteria and procedures for dealing with test results that may be inappropriate for statistical evaluation, due to for example, gross errors, miscalculations and transpositions.

ILAC G13 also discusses homogeneity and stability in many places in addition to the above, regarding competence of laboratory testing (essentially requiring ISO 17025 compliance for this), the quality of samples, planning the scheme, and in reporting of homogeneity and stability test results. For some accrediting bodies, if the laboratory carrying out the testing is not accredited to ISO 17025 for this work, they are subject to 2nd party audit by the PT provider for only the homogeneity/stability tests against ISO 17025. This approach requires that the provider have competent staff to carry out this audit (i.e. they must be trained and qualified as assessors or internal auditors).

ISO Guide 35 (2006)

5.8 …*Even when a material is expected to be homogeneous, as in the case of solutions, an assessment of the between-bottle inhomogeneity is required. When dealing with solid-state reference materials, including slurries and sludges, a within-bottle homogeneity study should be foreseen to determine the minimum sample intake…The minimum number of bottles selected at random is between 10 and 30, but should generally not be smaller than 10.*

Annex 2: Situations where homogeneity testing can be avoided.

2a: Use of inherent sample characteristics and process control

Introduction

There are clear requirements for homogeneity testing of samples in Standards and other reference documents for PT operation. These requirements stem from a recognition that samples delivered to participants should be sufficiently similar that differences between them will not affect the assessment of laboratory performance. For some PT samples, homogeneity testing of the samples is probably the only way to achieve this objective but in other cases there are more efficient means. This note identifies some cases where homogeneity testing will only add cost to PT operation in return for limited or no added value.

The Standards and reference documents also have to be suitable for all types of PT operation ranging from one-off exercises to frequent and regular distributions of similar samples. In the cases of the latter type of operation it is argued that sample production "process control" provides a more efficient means of quality control than sample testing of each batch. This approach to product quality control is the preferred one for manufacturing processes in general and this is exactly the type of operation involved in regular PT sample production.

True solutions

It is implicitly accepted in science that true solutions are homogeneous. For example, it is not considered necessary to take replicate aliquots of a master standard solution in the preparation of calibration standards to check that the master solution is homogeneous. So, why should PT samples that are solutions be subjected to sample testing? For PT samples there, of course, should be special care devoted to selection of solvents, and checks that a true solution has been generated with control mixing at all stages through process control. Also control is required of sample containers and dispensing, and of sample storage times and conditions. These extra aspects all form part of process control which has to be designed using knowledge and experience of the sample type and characteristics. Some examples are discussed in more detail below.

Individual sample preparation

There are some cases of samples that are individually prepared for example by adding a measured amount of one material to a measured amount of another. The intention may be to subject the entire sample to the analytical process so the homogeneity of the samples is determined by the uncertainties in the measurements of the two quantities. Sample testing will not add any value to these PT products; the necessary control is all in the sample preparation process.

Process control

PT schemes that run over many sample distributions can institute sample preparation process control as a means of ensuring the sample produced are fit for purpose in terms of their homogeneity. The stages involved could be as follows

1. Detailed specification of the sample preparation process including stages to be followed to ensure homogeneity of samples
2. Specification for selection and preparation of sample containers
3. Specification of the raw materials used
4. Process quality control charts to ensure procedures are within control limits
5. Established track record of laboratory performance from historic records of these specific samples or samples prepared using the same procedures
6. If these are new samples detailed technical consideration of the known physico-chemical properties of the substances involved is required
7. Comparison of each the performance of the samples from participants results with historic performance
8. Periodic review of overall performance as part of the scheme review process
9. Feedback to improve sample prep procedures

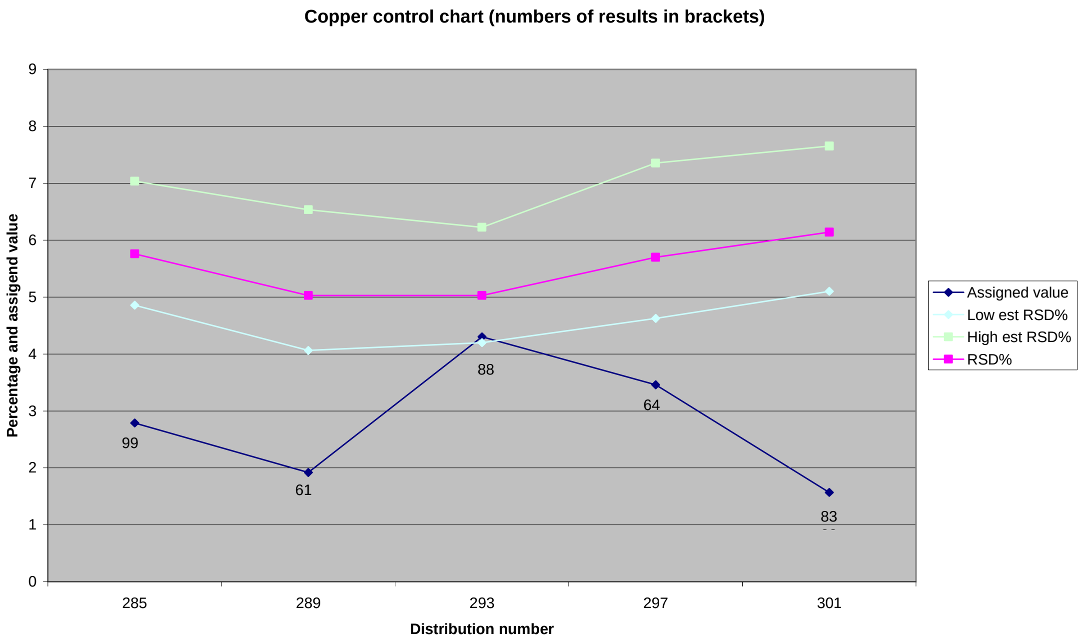
Example 1 – Augmented real water sample

A natural water sample is collected and stored for a period to allow for particulate settlement and other changes such as denitrification to proceed. The sample is then filtered for the preparation of the samples, additional salts are added as a solution, the mixture is stored for several days with regular mixing before being bottled and sent to PT participants. Results from all participants are as follows for chloride in five successive distributions:

Dist Number	No Results	Assigned Value	Robust RSD%
296	15	144	3.90
270	13	319	4.14
274	11	147	3.96
278	18	175	2.91
282	15	244	2.74

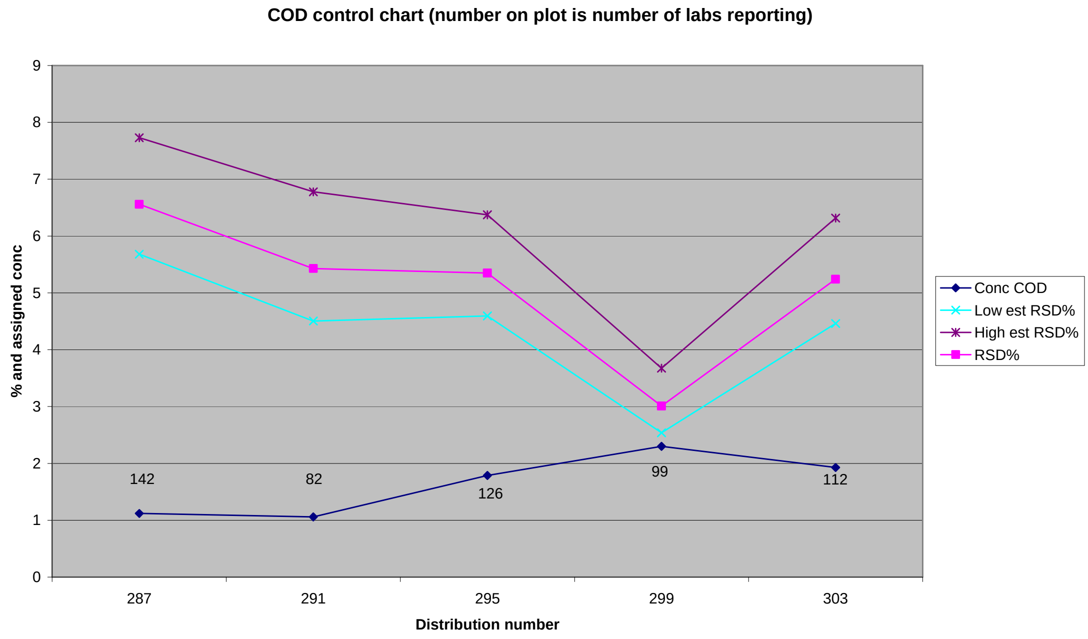
Two practical issues are worth mentioning. The first is that this is a fairly specialised product designed for those dealing with potable waters of quite high salinity, so it is a not a large market, hence the rather low number of participants. If homogeneity testing were undertaken it would involve a comparable number of samples to the number dispatched hence costs would be enhanced considerably, even without the analytical and organisational costs involved in the testing. Second the target error threshold for this determinand is 10% of the assigned value and the homogeneity target is generally set at 30% of this value i.e. in this example, a homogeneity target for the relative standard deviation (RSD) of 3%. Because chloride is fairly easy to analyse, results from participants are, in the last two distributions, within the homogeneity target! This can only happen if the samples are adequately homogeneous. However, evidence of this kind is only possible when the analytical methods used have very good performance.

The same procedure is used to prepare other samples and the plots below show the control chart for copper in surface water. Participants are free to use any analytical method and the analytical performance for copper is not expected to be as good as for chloride and this accounts for the larger RSD% values seen. The high and low estimates are 99%ile estimates. There have been approximately 70 previous distributions of similar samples so there is a considerable track record of performance available.



Example 2 – A spiking solution

A spiking solution of an organic chemical is used in a PT scheme for Chemical Oxygen Demand (COD). The solution is diluted by the receiving lab using deionised water following the instructions of the PT organiser and then subjected to COD analysis using the method of choice of the laboratory. The spiking solution sample is prepared in large volumetric glassware following essentially the stages in Standard methods for preparing standard solutions for COD analyses, with suitable checks on the methods followed. The control chart showing results from recent distributions is shown below



2b. Use of Participant Data for determining homogeneity and stability

Problems with homogeneity or instability should be reflected in the participant results. Inhomogeneous samples should be reflected in larger SDs for the laboratories or a higher rate of unacceptable results. For qualitative analytes this will appear as higher than usual numbers of incorrect results. Therefore, one possible alternative to homogeneity and stability testing is the analysis of participant data for indications of problems with the lot of samples. For the purposes of this discussion, samples affected by inhomogeneity or instability will be called "corrupted".

It can be expected that a corrupted lot of samples will produce results with poorer agreement than will an uncorrupted lot, especially when the problem affects only a portion of the samples in the lot. This would be true for both qualitative and quantitative tests, in any type of material.

A recent paper (Tholen, Chappel, Francis, Accreditation and Quality Assurance, Vol. 11, No. 8-9 / August, 2006, pp. 400-407.) proposes that for some qualitative External Quality Assessment (EQA) schemes, corrupted lots of proficiency testing samples can be detected by examination of participant data. This is in contrast to the usual examination with homogeneity tests prior to distribution and stability tests afterwards.

This proposition is examined initially for qualitative schemes only, and only if the following conditions apply:

- Similar samples have proven to be homogeneous and stable in past rounds;
- Rates of unacceptable results are consistent across rounds (follows an assumed statistical distribution);
- Participating laboratories are largely the same from round to round; and
- The processes for producing the samples are properly controlled, as evidenced by data, by accreditation, or by certification of the processes that produce the samples.

A statistical model was proposed to review the conditions under which homogeneity and stability testing need not be performed, unless a problem is indicated by participant data.

It is possible to develop decision rules for determination of adequate sample lot homogeneity and stability, based on participant results. The statistical model demonstrates that this procedure provides superior ability to detect problems for most situations of corruption.

For quantitative schemes, statistical rules based on PT data would be based on rates of unacceptable results (as for qualitative schemes), or on the between laboratory SD. With quantitative schemes, inhomogeneity might also appear as a bimodal distribution of results (depending on the nature of the inhomogeneity). Instability would appear perhaps as a trend of results changing by day of measurement.

To use this approach, the PT provider would need to have a statistical protocol specifically designed to take into account any inhomogeneity indicated by the participants' results, as well as needing to identify this amongst the normal spread of results which may be expected from a proficiency test. This approach could only be conducted if the following conditions are met:

a. An experienced, consistent set of laboratories from round to round;
b. Samples of demonstrated consistent quality in previous rounds;
c. The same or similar range of concentrations and matrices from round to round.
d. Data exist for at least 20 rounds that cover all concentration levels, and a demonstration of consistent SD's and/or rates of unacceptable results.

References

1. ISO/IEC Guide 43-1: Proficiency testing by interlaboratory comparisons – Part 1: Development and operation of proficiency testing schemes, 2nd Edition, 1997 International Organization for Standardization, Geneva.

2. ILAC G13:2007 "Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes." International Laboratory Accreditation Cooperation, Sydney, Australia.

3. Thompson M, Wood R. "The International Harmonized Protocol for the Proficiency Testing of (Chemical) Analytical Laboratories (Technical Report)". *Pure Appl Chemistry*, 65 (9): 2123-2144, 1993.

4. International Organization for Standardization, Geneva. ISO/IEC13528: 2005 Statistical methods for use in proficiency testing by interlaboratory comparison.

5. Thompson M, Ellison SLR, Wood R. "The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories" (IUPAC Technical Report) *Pure Appl Chem* (78) 1:145-196, 2006

6. Fearn T, Thompson M., "A new test for 'sufficient homogeneity'", *Analyst* 126, 1414-1417 (2001)

7 Korth W, Ralston J, Maddalena S, Connolly J. "The Role of proficiency testing in Australia's national residue Monitoring System". JACQUAL, in press, 2006

8. Korth W, Maddalena S, Roberts G, Briggs P, "Development, Implementation, and Summary Results of a Nitrofuran metabolites in Honey Programme" JACQUAL in press, 2006

9. ISO Guide 35: Reference materials – General and statistical principles for certification, 3rd Edition, 2006 International Organization for Standardization, Geneva.