THE DRINKING WATER INSPECTORATE

GUIDANCE ON

SAMPLE AND SAMPLE EXTRACT STABILITY TRIALS

DWI GUIDANCE ON SAMPLE AND SAMPLE EXTRACT STABILITY TRIALS

Purpose

The purpose of these trials is to demonstrate that the maximum permitted delay between sampling and analysis does not result in "a material alteration in the concentration or value for the measurement or observation of which the sample is intended" (regulation 16(2)(c)).

In terms of a statistical trial, the hypothesis to be tested is that the change in the mean of repeated measurements before and after storage is not greater than a target figure.

A successful stability trial must be conducted prior to adopting sample preservation conditions or sample extract preservation conditions which differ from, or storage times which are longer than, those documented in ISO 5667 Part 3:2003, or other authoritative source ie a standard method. Each sample matrix of interest must be fully tested. In practice, only the worst case or worst cases need be tested. For example testing nitrite stability in waters with low colony counts will not yield useful information.

Results of trials carried out in other laboratories may be used, provided the laboratory can show that the preservation and storage conditions are identical (not just similar) and that the sample matrices of interest to the laboratory were included in the trial and the results of the trial were fully satisfactory and robust (ie reproduced in three or more independent laboratories).

Specification of requirement

For regulatory analysis the appropriate target value is one half of the maximum permitted trueness error. For most parameters this is 5% of the value at the PCV. For many organic parameters it is 12.5%. Significance is at the two-sided 95% confidence level, and the power is set at 90%. Each sample matrix type of interest must be tested separately.

The specification, design, calculation and interpretation given in this document are all derived from NS30 pages 113 to 120, 137 to 139 and 148.

Design of Trial

The trial should consist of spiking of a pre-determined number of samples to the PCV. All samples must be collected by filling a series of bottles from the same source (eg a single tap). The true concentrations of the parameter must show negligible variation from one sample to another. If filling a series of bottles directly from the tap may not yield such samples, a bulk sample should be taken which is then mixed and sub-divided into a series of bottles. Precision of spiking is of paramount importance and more important than the absolute value spiked. If precision of spiking is likely to cause problems, consideration should be given to spiking a bulk sample, which can then be sub-divided into a series of bottles.

One set of samples is analysed on day 0, with a further set analysed at each selected time interval with all sample preservation and storage conditions applied exactly as it is intended to apply them to regulatory samples. It would be prudent to also include times less than the full period desired for routine storage of samples in the trial.
The estimated minimum number of samples (n) required to be analysed on each day of testing to show whether the change is significant is given below

Standard deviation (%PCV)	Number of samples to detect 12.5% change	Number of samples to detect 10% change	Number of samples to detect 5% change
1	2	2	2
2	2	2	5
3	2	3	10
4	3	7	17
5	5	**7** (mercury)	**26**
6	6	10	38
7	9	13	51
8	11	17	67
9	14	22	85
10	17	**26 (**tetrachloromethane)	104
11	21	32	126
12	24	38	150
12.5	**26**	41	163
38	38	59	234
20	67	104	416

These figures are minimum values of n for which the equation $(t_\alpha + t_\beta)\, s\sqrt{(2/n)} \leq \delta$ is true, where δ is the target change, subject to a minimum of 2 for a statistical comparison to be made. This indicates that the test will probably be sufficiently powerful to identify the target change as being a statistically significant difference. Figures in bold relate to the maximum permitted precision relevant to the maximum permitted change. These numbers are only estimates of the actual numbers required because the actual distribution of data will not be known until after the test is completed, and either more or fewer replicates may be needed in practice. Reasons for large deviations from the expected standard deviation should be investigated to determine if there is any reason for the unexpected change in performance, which may invalidate the trial. Large within batch variations can also lead to wrong conclusions being drawn. Prior to undertaking trials steps should be taken to ensure that between batch errors are not significant. The most common cause of significant between batch errors is variation in the true value of calibration standards. If it is not possible to reduce such errors to a magnitude which will not adversely affect the trial, means should be adopted to measure and compensate for such errors, such as those described in NS30 or DD ISO ENV 13530:1998

The same design can also be used to test alternative preservation and pre-treatment methods. In these cases, storage times should be the same and between batch errors can be eliminated by analysing both sets of samples in the same analytical batch.

Calculation

The significance of any observed difference is determined using a t-test. The following is an example calculation, with expected standard deviation of 2% and target change 5%

	Day zero	Day x
	101.0	94.0
	100.3	93.2
	98.8	92.9
	101.2	96.5
	99.9	92.8
Mean	100.225	94.05
Standard deviation	1.11	1.72

Pooled standard deviation	1.45
Mean difference	6.175
Standard Error (of differneces)	1.024
t statistic (calc)	**6.032**
Degrees of freedom	6
Critical value (.05) (from tables)	**2.447**

Conclusion: there is a real difference between the means.

Interpretation

$t_{0.05}$ for 6 degrees of freedom = 2.447 (from tables). The observed value is greater than the tabulated value and therefore there is a real difference between the two means. The numerical value of the change is also greater than the target value and therefore there is a significant change.

If the observed value of t is greater than the tabulated value and the change was less than or equal to the target change, the change is less than (or equal to) the target and samples may be stored for up to the tested period under the conditions tested.

If the observed value of t is less than the tabulated value and the change was equal to or greater than the target change, the trial was not sufficiently powerful to show a significant change and must be repeated with more replicates.

If the change is less than the target change and the observed value of t was less than the tabulated value then, provided the trial was sufficiently powerful and would have identified any difference in excess of the target change as being significant, there has been no significant change and samples may be stored for up to the tested period under the conditions tested. The test is sufficiently powerful if the target change is substituted for mean difference in the formula for the t test and the value of t then calculated is greater than the tabulated value. If it is not greater then the trial was not sufficiently powerful to show a significant change and must be repeated with more replicates.

In summary:

Mean difference greater than target?	Observed t greater than tabulated value?	Would difference equal to target change have observed t greater than tabulated value?	Proposed new storage arrangements satisfactory?
Yes	Yes	N/A	No
Yes	No	N/A	No*
No	Yes	N/A	Yes
No	No	No	No*
No	No	Yes	Yes

* Trial not sufficiently powerful to test the original hypothesis. Repeat trial using more replicates.