**Replicate analysis during analytical testing**

***Summary Document***

***Findings from a facilitated discussion, between peer laboratories, on replicate analysis.***

***Eurachem Ireland would like to thank all those who participated in the facilitated discussion.***

**Introduction**

The question of replicate analysis is one that must be considered when analytical testing is carried out. While guidance is available on the use of replicates during method validation [1, 2] as part of internal quality control [3] or the impact of replicate analysis on measurement uncertainty [4] and its use in compliance assessment [5], there is a lack of general guidance available on replicate analysis during routine use of a test method. Guidelines produced by the European Directorate for the Quality of Medicines and HealthCare (EDQM) are available which provide a protocol to manage pharmaceutical samples which are out-of-specification [11].

When analysing a test sample, how many replicates should be carried out? If the test sample must undergo repeat analysis, how many replicates should then be carried out? Should a laboratory have criteria in place for comparability of replicates and how are these criteria determined? What should be done if these criteria are not met?

Eurachem Ireland organised a facilitated discussion on replicate analysis in September 2021 that brought representatives of a number of Irish laboratories together to discuss and to share knowledge and their approaches to the management of replicate analysis. This document has been prepared based on the facilitated discussion and aims to provide some guidance to laboratories determining their own approach to replicate analysis during analytical testing.

**Procedure for replicate analysis**

Laboratories that carry out analytical testing should document a procedure for replicate analysis. When developing this procedure they should take into account legislation and sectoral guidelines which may have requirements for the number of replicates which must be carried out and what is reported to the customer. In the absence of such guidelines, assessing what the test method is used for and the needs of the customer should ensure the procedure for replicate analysis is fit for purpose.

The following are different approaches employed by the participating laboratories and it is suggested that they are points for consideration when a laboratory is devising its own policy:

*Initial analysis carried out singly*

* Initial analysis is a screening analysis. Any positive results will be repeated using a confirmatory method;
* Initial analysis is confirmatory but the analytes are banned. A negative result is likely and the aim of the initial analysis is to confirm a negative. Any positive results will be repeated;
* Sample size is limited.

*Initial analysis carried out in duplicate*

* Positive results are expected. For example, the analysis may be to determine compliance to a declared value;
* The results of the analysis are expected to be used in court;
* Historically, there has been a lack of matrix homogeneity.

*Repeat analysis carried out singly*

* Analysis is labour intensive and low numbers are analysed per batch;
* Repeat analysis is being carried out due to a QC or some other failure during initial analysis. There is no indication from the initial analysis that the test sample may be non-compliant;
* Sample size is limited.

*Repeat analysis carried out in duplicate*

* Initial analysis indicated that the test sample may be non-compliant.

If initial or repeat analysis is to be carried out in duplicate, the laboratory must document in advance what constitutes an acceptable replicate result and what is reported to the customer. The laboratory may choose to implement a repeatability criterion. This can be an absolute value, for example:

*The difference between the results of the duplicate determinations must not exceed:*

*0.2% in absolute value for ash contents less than 10%*

*Absolute values are calculated as follows: % Crude Ash (higher result) -% Crude Ash (lower result);*

or a relative value, for example:

*The difference between the results of the duplicate determinations must not exceed:*

*2% of the higher figure for ash contents of not less than 10%*

*Relative values are calculated as follows:*

*[(% Crude Ash (higher result) -% Crude Ash (lower result))/ Crude Ash (higher result)] x 100*

Repeatability criteria can be specified in legislation, sectoral guidelines or in EN or ISO standard methods. Alternately, repeatability criteria can be determined during validation of a method or by gathering data of replicate analysis during routine use of a method. Repeatability can be calculated using the standard deviation of the normalised differences. This procedure is described in the Eurachem/CITAC guide: Quantifying Uncertainty in Analytical Measurement [4].

The laboratory should document what will be reported to the customer. Usually, the average of the replicates is reported, once the repeatability and other acceptance criteria are met.

**Procedure when the acceptance criteria for replicate analysis are not met**

The laboratory’s procedure on replicate analysis should document the steps the laboratory follows if the acceptance criteria for replicate analysis are not met. This procedure will often involve further repeat analysis in duplicate. This can be an issue if the sample submitted to the laboratory is small. The laboratory may have insufficient sample to carry out another analysis. In this situation, the laboratory may consider implementing a minimum sample size and any samples smaller than this minimum are rejected on submission to the laboratory. Alternately, the laboratory may contact the customer and request further sample.

A failure of the replicate analysis to meet acceptance criteria may indicate a lack of homogeneity in the sample. Laboratories should consider the following when developing their replicate analysis procedure:

* Ensuring a representative laboratory sample is taken from the batch/lot being sampled is not the responsibility of the testing laboratory when the testing laboratory does not carry out the sampling. The testing laboratory can give relevant feedback to the body responsible for the sampling, however;
* A failure of the replicate analysis to meet acceptance criteria may indicate that the sample preparation steps such as mixing, milling, homogenisation etc. were insufficient. The procedure may include a re-homogenisation of the entire original sample before test aliquots are taken for repeat analysis;
* Certain sample preparation steps can cause samples and/or analytes to degrade and so these steps may only be carried out for repeat analysis;
* Freeze-thawing may impact sample and/or analyte stability;
* When laboratories are responsible for sampling, repeat analysis may include re-sampling.

**Reporting results when the acceptance criteria for replicate analysis are not met**

There are a number of approaches a laboratory may take when it has analysed a sample, repeated it in duplicate, repeated in duplicate again following re-sampling/re-homogenising if appropriate, and the replicate analysis still fails to meet acceptance criteria. Whichever approach a laboratory chooses to take, it should be agreed with the customer and documented in the replicate analysis procedure. If the test method is accredited, it may be advisable to discuss the approach with INAB. In each case it should be noted on the test report that the replicate analysis failed to meet acceptance criteria. Some suggested approaches are:

* The average of all replicates is reported;
* The average of the repeat analysis replicates is reported;
* The average of replicates which are within the MU of the test method of each other is reported;
* The average of the two closest replicates is reported;
* The lowest of the results is reported – this may be useful in areas where maximum levels are set in legislation and the legislation is written so as to give the “benefit of the doubt” to the producer;
* The average of replicates may be reported as an “indicative” result or state that the result is “in the order of”;
* The laboratory may state that a test report cannot be issued. In this case, the customer should be contacted and the reason for not issuing a test report discussed.

**The use of outlier tests**

When a laboratory has analysed a test sample a number of times and repeatability has failed, the use of outlier tests may be considered to determine if one or more of the data points are outliers. Incorrectly retaining or omitting a data point can have a significant effect on a test result, particularly if the dataset is small. We must have a basis for data rejection, rather than just thinking a data point looks “out”. The ISO recommended test for outliers is the Grubbs test [6].

*Grubbs test for outliers*

* The Grubbs test is based on significance testing;
* It assumes the population has a normal error distribution;
* It tests the hypothesis that all measurements come from the same population;
* The G statistic is calculated (Gcal);
* The critical value of G (Gcrit) is obtained from statistical tables (significance level in this case is 0.05);
* If Gcal>Gcrit, then the null hypothesis (H0) is rejected and the suspect value is deemed to be an outlier.

Calculation of G statistic:

Note: and *s* are calculated with the suspect value included, as H0 presumes that there are no outliers.

*Q-Test (Dixon’s test)*

Another test which may be used to determine if a suspect value is an outlier is the Q-test.

* The Q-test compares the differences between a suspect value and the measurement nearest to it in size within the range of measurements.
* It assumes the population has a normal error distribution;
* It tests the hypothesis that all measurements come from the same population;
* The Q statistic is calculated (Qcal);
* The critical value of Q (Qcrit) is obtained from statistical tables (significance level in this case is 0.05);
* If Qcal>Qcrit, then H0 is rejected and the suspect value is deemed to be an outlier.

Calculation of Q statistic:

*Usefulness of outlier tests for replicate analysis – the power of n*

The usefulness of the outlier tests in replicate analysis can be demonstrated using the following example.

The following values were obtained for the nitrite concentration (mg/L) in a sample of river water:

0.403, 0.410, 0.401, 0.38.

The last measurement is noticeably lower than the others and is suspected of being an outlier. Should it be rejected from the data set?

Both the Grubbs and the Q tests for outliers were applied to the data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Calculation | Test Statistic | Critical Value | Outlier? |
| Grubbs | (0.380-0.399)/0.013 | 1.46 (Gcal) | 1.481 (Gcrit) | No |
| Q | (0.380-0.401)/(0.410-0.380) | 0.7 (Qcal) | 0.829 (Qcrit) | No |

The result of both outlier tests agree – there is no evidence to suggest that the suspect data point can be designated as an outlier and therefore its removal from the dataset cannot be justified.

Three further measurements were made so that the complete dataset (n=7) is:

0.403, 0.410, 0.401, 0.380, 0.405, 0.413, 0.408.

Again, both the Grubbs test and the Q-test were used to determine if 0.380 is an outlier.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Calculation | Test Statistic | Critical Value | Outlier? |
| Grubbs | (0.380-0.403)/0.011 | 2.09 (Gcal) | 2.02 (Gcrit) | Yes |
| Q | (0.380-0.401)/(0.413-0.380) | 0.636 (Qcal) | 0.568 (Qcrit) | Yes |

This time, there was sufficient evidence to classify 0.380 as an outlier and therefore warrant its removal from the dataset.

This highlights that the ability of outlier tests to classify outliers increases as the number of data points in the replicate dataset increases. However, as n increases, the impact that a single outlier will have on the mean value decreases. Therefore, as datasets from replicate analysis tend to be quite small, the usefulness of outlier tests may be limited. In addition, the presence of more than one outlier can result in masking effects. Handling multiple outliers is possible but increases complexity of analysis. For the Grubbs test, the minimum sample size advised is 7. While the Q test is recommended for smaller sample sizes, the tests are increasingly reliable with higher sample sizes.

**Conclusion**

Laboratories carrying out analytical testing should have a documented procedure for replicate analysis. The procedure should record how many replicates are carried out for initial analysis of a test sample and how many are carried out for repeat analysis. This should be based on legislation, sectoral guidelines, what the test method is to be used for and the needs of the customer. If replicate analysis is carried out, the procedure should include acceptance criteria for the replicates. The procedure should also document the steps carried out if replicate analysis fails to meet acceptance criteria. When determining what should be reported to the customer, outlier tests can be a useful tool. They must be treated with caution as they assume a normal distribution, are less powerful with a small dataset and can be impacted by the presence of more than one outlier. The procedure for replicate analysis should document if a test report will be issued to the customer when replicate analysis fails to meet acceptance criteria and, if a test report is issued, how the result will be reported.

**References and further reading**

|  |  |
| --- | --- |
| 1 | B. Magnusson and U. Örnemark (eds.) Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, (2nd ed. 2014). ISBN 978-91-87461-59-0. Available from http://www.eurachem.org |
| 2 | V. Barwick (ed.), Planning and Reporting Method Validation Studies – Supplement to Eurachem Guide on the Fitness for Purpose of Analytical Methods (2019). Available from http://www.eurachem.org |
| 3 | V. Barwick (Ed), Eurachem/CITAC Guide: Guide to Quality in Analytical Chemistry: An Aid to Accreditation (3rd ed. 2016). ISBN 978-0-948926-32-7. Available from www.eurachem.org. |
| 4 | S L R Ellison and A Williams (Eds). Eurachem/CITAC guide: Quantifying Uncertainty in Analytical Measurement, Third edition, (2012) ISBN 978-0-948926-30-3. Available from www.eurachem.org |
| 5 | A. Williams and B. Magnusson (eds.) Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment (2nd ed. 2021). ISBN 978-0-948926-38-9. Available from www.eurachem.org. |
| 6 | Barnett, V. and Lewis, T., 1984, Outliers in Statistical Data, 2nd edn, John Wiley & Sons Limited. |
| 7 | Evaluation and reporting of results, OMCL Network of the Council of Europe, 2014. |
| 8 | Statistics for the Quality Control Chemistry Laboratory, E. Mullins RS.C (ISBN 0-85404-671-2) |
| 9 | Nordtest Report TR 569 (ed 5.1) Internal Quality Control |
| 10 | SOFT/AAFS Forensic Toxicology Lab Guidelines |
| 11 | Guidelines produced by The European Directorate for the Quality of Medicines and HealthCare (EDQM). They are set up to deal with samples which are out-of-specification:  [https://www.edqm.eu/sites/default/files/medias/fichiers/OMCL\_Network/Find\_information\_on/Quality\_Management\_Documents/omcl\_evaluation\_and\_reporting\_of\_results\_-\_core\_document.pdf](https://scanner.topsec.com/?r=show&u=https%3A%2F%2Fwww.edqm.eu%2Fsites%2Fdefault%2Ffiles%2Fmedias%2Ffichiers%2FOMCL_Network%2FFind_information_on%2FQuality_Management_Documents%2Fomcl_evaluation_and_reporting_of_results_-_core_document.pdf&d=2120&t=4344ac775fa9c853f5d9c56003213f12e90455c8)  One of the annexes deals with reporting for quantitative tests:  [https://www.edqm.eu/sites/default/files/medias/fichiers/OMCL\_Network/Find\_information\_on/Quality\_Management\_Documents/omcl\_evaluation\_reporting\_of\_results\_annex\_2a\_re\_test\_programmes\_for\_qualitative\_tests.pdf](https://scanner.topsec.com/?r=show&u=https%3A%2F%2Fwww.edqm.eu%2Fsites%2Fdefault%2Ffiles%2Fmedias%2Ffichiers%2FOMCL_Network%2FFind_information_on%2FQuality_Management_Documents%2Fomcl_evaluation_reporting_of_results_annex_2a_re_test_programmes_for_qualitative_tests.pdf&d=2120&t=6bdbfde0914502bd90e5b4613690cf17ce9415a2) |