DATA ANALYSIS AND PERFORMANCE ASSESSMENT

The purpose of External Quality Assessment is to compare performance of individual laboratories or test centres with their peers. EQA is a retrospective form of quality control – participants do not know the target range or expected results for their test, unlike internal quality control.

Target values: Performance is determined by comparison of individual laboratory results with the target value for each test. UK NEQAS BC uses **median values** determined from participants' results to determine the consensus or "target" values against which individual laboratory performance can be assessed. Use of the median avoids the effect of outlying results and the need to remove any outlying data from the analysis.

Peer Groups: Where consistent reagent or method-related differences have been identified, or may be expected, participants' results are assessed against their 'peer-groups' provided the number in that group is sufficient to be statistically valid.

Performance terminology: Results are assessed using statistical measures comparing individual results with peer group or overall consensus values. Therefore, a result falling outside the statistically defined terms is not necessarily a "poor" or "unsatisfactory" result. We deliberately employ the term ***"outwith consensus"*** to indicate a result that falls outside the statistically determined limits we have applied.

Performance Analysis: Two different approaches are taken to evaluate performance for screening tests and factor assays – both comply with specifications detailed in ISO 13528 *Statistical Methods for use in proficiency testing by interlaboratory comparisons.*

Screening Tests

For **INR, PT for diagnosis**, and **APTT** the percentage deviation of each individual laboratory's results from the reagent and overall medians are calculated and the following criteria for performance are applied:-

Performance is considered "*within consensus*" if the deviation is ≤15% from:

The **reagent median** if the number of users of that reagent is equal to or greater than 10
or
The **overall median** if the number of users of the reagent is less than 10.

Results >15% deviation from the median are considered "*outwith consensus.*"

For **thrombin times**, 20% deviation is used as the acceptable limit for "within consensus" performance.

For **Heparin Dosage Assessment (HDA) by APTT**, performance is considered "*within consensus*" if the deviation is ≤20% from the reagent median if the number of users of that reagent is equal to or greater than 10. Results >20% deviation from the median are

considered "*outwith consensus.*" Marked differences in the heparin sensitivity of APTT reagents have led to the conclusion that it is inappropriate to assess minority reagent users against the overall median. No performance analysis is applied to minority groups (n<10 users), although % deviation from reagent and overall medians are recorded on individual reports.

For Fibrinogen assay, Clauss method results are assessed against the overall Clauss method median, with results >15% from this median considered *"outwith consensus"*. Multifibrin U users are assessed separately. *EQA is not offered for PT-derived fibrinogen, as this assay is not recommended by expert groups.*

Factor Assays

For factor assays, UK NEQAS BC distributes samples with factor concentrations covering the wide range encountered in clinical practice. For this reason, percentage deviation from the median is inappropriate as a means of defining performance. A ranked grading analysis to evaluate performance was devised by Professor S Thomson, Department of Medical Statistics & Evaluation, Royal Postgraduate Medical School, London.

The overall consensus median is taken as the central reference point or "target value". Individual results are ranked into 5 unequal quantiles above and below the median, each quantile being designated by a letter depending on ranked distance from the median:

A: *The nearest 25% of results above (A) and below (a) the median (i.e. 50% of results);*

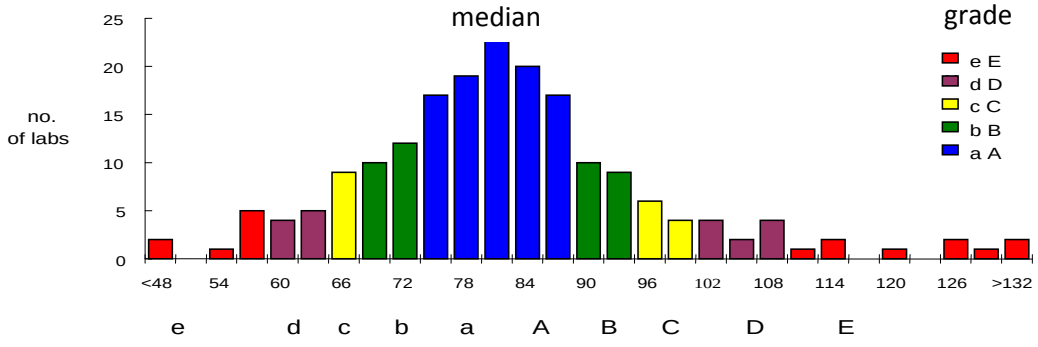
B: *The next 10% of results above (B) and below (b) the median (i.e. 20% of results);*

C: *The next 5% of results above (C) and below (c) the median (i.e. 10% of results);*

D: *The next 5% of results above (D) and below (d) the median (i.e. 10% of results);*

E: *The 5% of results furthest from the median, above (E) and below (e) (i.e. 10% of results).*

This is illustrated below:



Grades below the median are shown in lower case, and above the median in upper case, to aid in assessment of bias.

Performance is based on grades obtained in a minimum of **two consecutive exercises** for any particular test. ***"Outwith consensus"*** is defined as a combination of a C (or 'c') grade together with an E (or 'e') grade, or any combination of D (or 'd') and E (or 'e') grades (e.g. cE, ec, Dd, de, ED and EE in consecutive distributions of that particular assay).

Persistent "outwith consensus" performance is defined as two consecutive ***"outwith consensus"*** performances, where the order in which the grades were assigned does not affect the overall performance. This will arise from three consecutive performances with the following combinations of grades (upper case only shown):

DDD, DED, ECE, EEC, DDE, DEE, EDD, EED, CEE, EDE, EEE

A non-return for a registered test will be graded as 'F' and taken as equivalent to an E grading. Thus, designations which include 'F' grades are based on performance over 2 or 3 exercises, respectively.

Exceptions: In some cases, significant differences have been noted between different methodologies. Where this occurs on a consistent basis, separate analysis of the groups is carried out, using medians specific to each method group. However, the A-E grading system can only be applied is only effective if the number of participants is equal to or greater than 20; consequently, grading analysis is not applied to groups of results from fewer than 20 centres. In these cases, reports indicate performance is "not assessed" with a grade of "*" , and we recommend that the participant compares their result with the median for users of the same method alongside the overall assay median to evaluate whether any differences from these results have clinical relevance.. We recommend this approach for *all* participants, including those receiving a performance assessment of their results. Advice is always available by making contact with the Scheme.

The following assays are currently analysed in separate peer groups. These are regularly reviewed:

- D-Dimers Assays (kit-specific and FEU/non-FEU groups)
- Factor VIII:C assay (1-stage, 2-stage & chromogenic assays)
- Antithrombin antigen (results expressed in u/dl and mg/dl)
- Antithrombin activity (bovine thrombin, human thrombin, factor Xa substrate)
- Protein C activity (clotting and chromogenic assays)
- Activated Protein C resistance (Kit-specific groups)
- VWF RiCof (activity) (Kit-specific groups)

Performance analysis for protein S activity assays is currently suspended.

Consequence of outwith consensus performance

If results of screening tests are outwith consensus on three consecutive occasions (including failure to return results), or results from factor assays are persistently outwith consensus, a letter of concern with an offer of assistance is sent to the Head of Department by the Scheme Director. The Director will also make contact after a single survey for any results deemed to be sufficiently far from the target result to be considered clinically hazardous.