Performance Monitoring System for the Myeloproliferative Neoplasms Diagnostic Testing (MPN DT) Programme

Aim

The scoring system is a rolling scheme that will identify unsatisfactory performance or persistent unsatisfactory performance of any participant. This is in order that UK NEQAS LI can provide support and guidance where needed and ensure that the Genetics National Quality Assurance Advisory Panel (NQAAP) are informed as appropriate. Please note that each EQA programme will be scored independently.

Outline

Two samples are issued for each trial that may or may not feature a clone with a MPN driver mutation: either *JAK2* p.(Val617Phe), or a clinically significant variant in *JAK2* exon 12, *CALR* exon 9, or *MPL* exon 10. There are four trials per annum. Participants are asked to test the majority of samples for each of the four variant types, subject to their laboratory offering an appropriate assay, ie testing for all four variant types is not compulsory; testing should reflect a laboratory's standard test repertoire. Incorporated within the eight trial samples issued each year will be a variety of *JAK2* p.(Val617Phe) positive samples with differing allele burdens, including occasional samples that reflect measurable residual disease (MRD) post treatment / allogeneic haematopoietic stem cell transplant (aHSCT). These samples will be identified at trial issue as requiring post treatment testing and should only be tested for *JAK2* p.(Val617Phe).

Qualitative Results and Performance Monitoring

The MPN DT programme requires a qualitative response from participants for each of the four variant types, subject to that laboratory performing the relevant testing. Participants are asked if, using their normal laboratory technique, they detected the presence of:

- the *JAK2* p.(Val617Phe) variant
- a clinically significant variant in exon 12 of the *JAK2* gene
- a clinically significant variant in exon 9 of the *CALR* gene
- a clinically significant variant in exon 10 of the *MPL* gene

In line with BS EN ISO/IEC 17043:2023 standards, the presence/absence of each of these variants is determined from the consensus of all participants' results (modal result). Each participant response is then compared against the consensus results. If the participant is out-of-consensus for one or both samples a Critical (Unsatisfactory) status is awarded for that trial. Please note that if a participant does not provide one or more of the relevant tests (including MRD testing for the *JAK2* p.(Val617Phe) variant), they will have the option to inform UK NEQAS LI at data entry so that the laboratory can be excluded from scoring of relevant results.

Non returns will result in an immediate Critical status for that trial.

If a participant is awarded two or more Critical statuses out of three trials issued, then their overall status will escalate to Persistent Unsatisfactory Performance.

Please note, results should not be submitted if samples fail internal quality control (QC) measures. Repeat samples are available for all trials, if required. If following repeat sample(s) processing, results obtained still do not pass local internal QC please contact UK NEQAS LI. If results are submitted based on suboptimal results, they will be subjected to the same performance monitoring mechanisms as all other participants.

Additional Quantitative *JAK2* p.(Val617Phe) Results and Performance Monitoring (optional)

Participants also have the option to return quantitative results (variant allele burden) for samples in which the *JAK2* p.(Val617Phe) variant has been detected (variant allele burden (%) = 100x variant alleles / total alleles). A participant's result is then used in conjunction with the sample robust mean (RM) and robust standard deviation (RSD) to calculate a z score using the following formula:

$$z = (x - X)/\hat{\sigma}$$

where x is the result returned by the participant, X is the assigned value (RM) and $\hat{\sigma}$ is the standard deviation for proficiency testing (RSD).

The robust mean and robust SD are derived from participant data using Algorithm A (ISO 5725-5) which ensures that all data are included in the generation of the robust mean and robust SD but also minimizes the effect of outliers upon the final values. The robust mean and SD are calculated to 2 decimal places.

Interpretation of z scores in the context of this programme is as follows:

- A result between 2.5 and -2.5 indicates Satisfactory Performance.
- A result between 3.5 and 2.5 or -2.5 and -3.5 is considered as an 'Action' result, highlighting a potential issue to the laboratory.
- A result above 3.5 or below -3.5 is deemed to be a 'Critical' result requiring immediate investigation by the laboratory.

- Two 'Action' results in a period of any three consecutive samples would also be classed as a 'Critical' result. Please note that whilst each 'Action' is considered in combination with any other 'Actions' within rolling 3-sample periods, the same pair of 'Action' scores will not be combined to result in a 'Critical' classification more than once.

Due to the nature of how z scores are generated a positive z score highlights a positive bias in a laboratory's results whereas a negative z score shows a negative bias. Thus, the z score immediately highlights to the participating centre if their result is above or below the expected consensus value. Furthermore, the normalisation of z scores using the robust SD allows direct comparison of performance across multiple samples/trials.

Whilst participants can use their z score for information purposes only, we now also offer performance monitoring based on these z score classifications. This is currently optional and operates in addition to and independent of the obligatory performance monitoring of qualitative results. However, we strongly recommend that participants providing *JAK2* p.(Val617Phe) quantification on clinical reports participate in this additional layer of performance monitoring.

Unsatisfactory quantitative performance is defined as any Critical quantitative classification. If a participant amasses three Critical quantitative classifications within six relevant, consecutive, *JAK2* p.(Val716Phe) positive samples, their status is elevated to Persistent Unsatisfactory Performance for *JAK2* p.(Val716Phe) quantification.

Please note that qualitative and quantitative Critical scores are considered separately within the respective independent qualitative and quantitative performance monitoring systems; they are not considered together to elevate a participant to Persistent Unsatisfactory Performance.

Following Unsatisfactory Performance

Unsatisfactory Performance, for either qualitative and/or quantitative results, will be initially communicated to participants on their trial report. This will be followed up with an email and notification on the Participant Hub, highlighting that performance on the last sample(s) was out of consensus and offering support and guidance to assist in returning to Satisfactory Performance. The support and guidance offered will be tailored to the particular needs of the participant but may include the provision of repeat/additional samples, communications by email, telephone conversations or face-to-face communications.

If a participant's status is elevated to Persistent Unsatisfactory Performance, then a further email will be issued highlighting this and, for UK based laboratories, the Genetics NQAAP panel informed. Participant's results will be reviewed by the lead scientist and any UK participant may, at the discretion of the Director and Specialist Advisory Group chairperson, be referred Genetics NQAAP even if they have not met the criteria for Persistent Unsatisfactory Performance in any individual EQA.

As with all scoring systems it is important to note that these will be constantly reviewed to determine if they are providing the information required. The Director of the scheme retains the discretion to determine if any individual trial should not be scored.