UK NEQAS LI *BCR*::*ABL1* (major) Quantitation – Molecular Programme

Performance Scoring System

Aim

The performance monitoring system is a rolling scheme that will identify unsatisfactory performance or persistent unsatisfactory performance of any participant. This is in order that UK NEQAS for Leucocyte Immunophenotyping (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 programme is scored independently.

Measurement of *BCR*::*ABL1* is used to manage patient treatment protocols in chronic myeloid leukaemia (CML), and has been incorporated into international treatment recommendations (Hochhaus *et al*, 2020). In international treatment recommendations, molecular response in CML must be reported according to the International Scale (IS), as the ratio of *BCR*::*ABL1* transcripts to *ABL1* transcripts, or to other internationally accepted control transcripts (e.g., *GUSB*), and expressed and reported as *BCR*::*ABL1* % on a log scale (Hochhaus *et al*, 2020). As such a performance monitoring system has been developed that scores individual sample % *BCR*::*ABL1* ^{IS} results (Single Sample Scoring).

However, this performance monitoring system is only applicable laboratories using *ABL1* as a control gene as validation data suggested that alternative reference gene expression (e.g. *GUSB*, *BCR* and *B2M*) in the EQA material (compared to patient material) is significantly different and causing these labs to no longer be reporting on the IS. As the programme does not have a sufficient number of alternative reference gene users to robustly score them as individual user groups on single sample results, they will be scored based on the log change between two samples (see Log Change Scoring below). A small number of participants do not submit IS results and will also be scored based on the log change scoring system.

Outline

Standardly, two samples are issued in each trial with varying levels of % *BCR*::*ABL1* /reference gene. There are a maximum of 3 trials per annum. Participants are asked, using their normal laboratory technique, to report a % *BCR*::*ABL1*/reference gene result for each sample, unconverted or converted to the IS, to however many decimal places they would clinically. Participant's will then be scored based on either the Single Sample or Log Change methods based on the criteria outlined above.

Single Sample Scoring (IS results only)

Because any error in RT-qPCR is multiplicative, rather than additive, data distributions from RT-qPCR based EQA/PT testing programmes produce a lognormal distribution, i.e. an asymmetric distribution of results with a strong positive skew. To normalise the distribution and thus enable calculation of variation using parametric methods, log10 transformation of participant's % *BCR*::*ABL1* ^{IS} results is undertaken.

From the participant's log transformed % *BCR*::*ABL1*^{IS} results, a z-score is calculated using the following formula:

$$z = \frac{(log10x - log10x_a)}{\sigma_p}$$

where x is the participant's result

x_a is the assigned value (robust mean of log transformed data set)(ISO, 2015)

σ_p is the standard deviation of the proficiency test. This is a fitness for purpose criterion and has been set at 0.2386 which equates to a critical result being > threefold from the assigned value (that is a ratio of the assigned value to local result < $0.33x_a$ or >$3.0x_a$) in line with EUTOS criteria(Müller *et al*, 2009).

The robust mean is derived from participant data using Algorithm A (ISO 5725-5) that ensures that all data is included in the generation of the robust mean, but also minimises the effect of outliers upon the final value. The robust mean is calculated to 2 decimal place (d.p.).

This method of calculating z-scores has previously been established in the literature (Powell & Owen, 2002; Analytical Methods Committee, 2004; Thompson *et al*, 2006)

- A result between 2.0 and -2.0 will be classed as satisfactory
- A result between 3.0 and 2.0 or -2.0 and -3.0 is seen as an 'action' result. This highlights a potential issue to the laboratory. Two 'action' results in a period of 3 samples would result in classification as a 'critical'
- A result above 3.0 or below -3.0 is considered to be a 'critical' result requiring immediate investigation by the laboratory

If no *BCR*::*ABL1* is detected, in a sample where the consensus is that *BCR*::*ABL1* is present, this is classified as a Critical result.

Any laboratory who fails to return a result by the closing date will be classified as 'Action' for each sample. **Please note, results should not be submitted if samples fail internal quality control measures. Repeat samples are available for all trials, if required. Following repeat sample(s) processing, if results obtained still do not pass local internal QC please contact UK NEQAS LI. If results are submitted based on the suboptimal results they will be subjected to the same performance monitoring mechanisms as all other participants.**

Unsatisfactory performance for single sample scoring is defined as any occurrence of critical performance and this will be initially communicated to participants on their trial report. This will be followed up with an email and notification on the participant hub on each occurrence of unsatisfactory performance highlighting that performance on the last sample(s) was out of consensus and offering support and guidance to assist in returning to satisfactory performance. This may take the form of repeat/additional samples, communications by email, telephone conversations or face to face communications. If a participant amasses three critical performances within a 3 trial (6 sample) period their status is elevated to persistent unsatisfactory performance then a further letter will be issued and the Genetics National Quality Assurance Advisory Panel (NQAAP) informed (for UK participants only).

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. As such, this adds value to the performance monitoring information provided to laboratories because the z-score

immediately highlights to the participating centre if their result is above or below the expected consensus value.

Participant's results will be reviewed by the lead scientist and the participant may, at the discretion of the Director and Specialist Advisory Group chair person, 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 the limits will be constantly reviewed to determine whether they are providing the information required. The Director of the programme retains the right to determine if an individual trial should not be scored or scoring be amended.

Log Change Scoring

From the participant's submitted results for each sample a log change is calculated using the following formula:

$$log10(\frac{BCR :: ABL1/Reference\ Gene\ (\%)\ Sample\ 2}{BCR :: ABL1/Reference\ Gene\ (\%)\ Sample\ 1})$$

The log change value is then used in conjunction with the robust mean and robust standard deviation to calculate a z-score using the following formula:

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

where x is the result returned by the testing laboratory,
X is the assigned value (robust mean) and
$\hat{\sigma}$ is the standard deviation for proficiency assessment (robust SD).

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

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

- A result between 2.5 and -2.5 will be classed as satisfactory
- A result between 3.5 and 2.5 or -2.5 and -3.5 is seen as an 'action' result. This highlights a potential issue to the laboratory. Two 'action' results in a period of two trial issues would result in classification as a 'critical'
- A result above 3.5 or below -3.5 is considered to be a 'critical' result requiring immediate investigation by the laboratory.

For log change scoring, participants can choose to be scored on their unconverted % *BCR*::*ABL1*/reference gene, their % International Scale (%IS) results or both results depending on their laboratory preference. If a participant chooses to be scored on both the % International Scale (%IS) and unconverted % *BCR*::*ABL1*/reference gene submitted results, a maximum of one 'Action' or 'Critical' result is used per trial distribution to inform running performance.

If no *BCR*::*ABL1* is detected in either sample where the consensus is that *BCR*::*ABL1* is present this is classified as a Critical result.

Any laboratory who fails to return a result by the closing date will be classified as 'Critical' for the trial**. Please note, results should not be submitted if samples fail internal quality control measures. Repeat samples are available for all trials, if required. Following repeat sample(s) processing, if results obtained still do not pass local internal QC please contact UK NEQAS LI. If results are submitted based on the suboptimal results they will be subjected to the same performance monitoring mechanisms as all other participants.**

Unsatisfactory performance in this programme is defined as any occurrence of 'Critical' performance and this will be initially communicated to participants on their trial report. This will be followed up with an email and notification on the participant hub on each occurrence of unsatisfactory performance highlighting that performance on the last sample(s) was out of consensus and offering support and guidance to assist in returning to satisfactory performance. This may take the form 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 (defined as a 'Critical' classification on 2 or more occasions within a 12-month period) then a further letter will be issued and the Genetics NQAAP informed (for UK participants only).

Participant's results will be reviewed by the lead scientist and the participant may, at the discretion of the Director and Specialist Advisory Group chair person, 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 the limits will be constantly reviewed to determine whether they are providing the information required. The Director of the programme retains the right to determine if an individual trial should not be scored or scoring be amended.

Reference(s)

Analytical Methods Committee (2004) GMO Proficiency testing: Interpreting z-scores derived from log-transformed data. *The Royal Society of Chemistry***,** 1–2.

Hochhaus, A., Baccarani, M., Silver, R.T., Schiffer, C., Apperley, J.F., Cervantes, F., Clark, R.E., Cortes, J.E., Deininger, M.W., Guilhot, F., Hjorth-Hansen, H., Hughes, T.P., Janssen, J.J.W.M., Kantarjian, H.M., Kim, D.W., Larson, R.A., Lipton, J.H., Mahon, F.X., Mayer, J., Nicolini, F., et al (2020) European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia*, **34,** 966–984.

ISO (2015)ISO 13528:2015. Statistical methods for use in proficiency testing by interlaboratory comparison.

Müller, M.C., Cross, N.C.P., Erben, P., Schenk, T., Hanfstein, B., Ernst, T., Hehlmann, R., Branford, S., Saglio, G. & Hochhaus, A. (2009) Harmonization of molecular monitoring of CML therapy in Europe. *Leukemia*, **23,** 1957–1963.

Powell, J. & Owen, L. (2002) Reliability of food measurements: the application of proficiency testing to GMO analysis. *Accreditation and Quality Assurance*, **7,** 392–402.

Thompson, M., Ellison, S.L.R., Owen, L., Mathieson, K., Powell, J., Key, P., Wood, R. & Damant, A.P. (2006) Scoring in genetically modified organism proficiency tests based on log-transformed results. *Journal of AOAC International*, **89,** 232–239.