UK NEQAS
Immunocytochemistry & In-Situ Hybridisation

5 Coldbath Square, London EC1R 5HL
Phone: +44(0)208 187 9174
Email: info@ukneqasiccish.org

Ki-67 in Breast Cancer Module

Guidance Notes to Aid in the Interpretation of the Quantitative Results

Background Information

These notes are intended to provide additional information about your quantitative Ki-67 staining results to help you interpret them and guide you in deciding if any action is required.

The quantitative findings are supplementary to your standard qualitative report.

At this time, they are not being used in the performance monitoring process.

They have been produced using proprietary RUO software (Visiopharm: https://visiopharm.com/app-center/app/ki-67-breast-cancer/).

We have carried out quantitative analysis of the slides submitted by the whole cohort of participants, and the descriptive statistics that are produced are presented without any outlier trimming, or selection on the basis of qualitative score achieved.

The normal distribution

As is to be expected, the frequency of quantitative scores that are derived from the submitted stained slides at any given assessment run follow a normal, or Gaussian, distribution (see Figure 1). Albeit one that shows a degree of 'skew'.

In an ideal normal distribution data is dispersed around the mean (which coincides with the median and the mode), with this being the point at which the highest frequency of data-points occurs. Distribution frequency decreases with distance from the mean with half of the values above the mean and half below, resulting in a characteristically shaped curve termed the 'bell curve' (see Figure 1).

The standard deviation (SD) reflects the dispersion, or variability of the distribution. A large SD indicates a wider 'spread' of results, while a low SD indicates a more tightly group distribution.

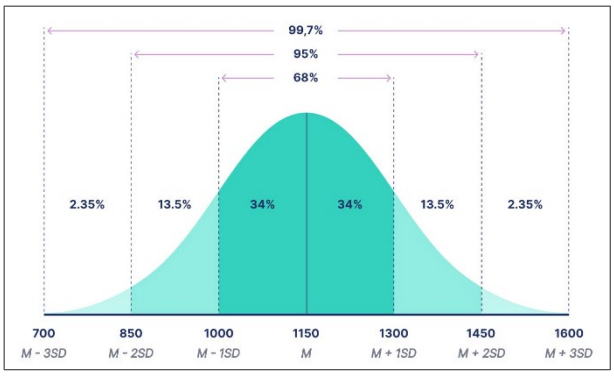

Figure 1 illustrating how the mean and SD describe a normal distribution curve.
M = mean, SD = standard deviation. Arbitrary values have been used on the x-axis.

Confidence intervals

The confidence interval (CI) gives the probability that a result will fall within a pair of values distributed around the mean, thus measuring the degree of certainty in a sampling method.

If several random samples were collected, the mean for that variable would differ slightly from one sample to another. Therefore, instead of providing only one value, a range of values (an interval) are specified within which this mean is likely to be located. To obtain a CI the margin of error is either added (upper CI) or subtracted (lower CI) from the sample mean. The range will be wider or narrower depending on the degree of certainty. By convention the 95% CIs are used.

We have used the range of scores within the 95% CIs to define the acceptable score-range. This is in-line with practice in a wide variety of EQA schemes that produce results in the form of quantitative data.

Result Interpretation

At each assessment run you receive three sets of results, one for each UK NEQAS-provided tissue. The

tissues we have provided give the most useful information in the context of the breast cancer clinical setting.

These are:

- Reactive tonsil (the DIA app we use assesses germinal centres selectively),
- An invasive breast cancer sample showing a low-level of proliferation (typically <10%).
- An invasive breast cancer showing a moderate to high-level of proliferation (typically >20%).

For each sample you are provided with the mean and the SD for the frequency distribution. The 95% CIs are also given together with an indication about where your result falls in relation to those CIs.

If your quantitative results fall within the 95% CIs for the distribution, no action is required. We would regard this as the 'ideal' result.

If your result falls outside the 95% CIs and is less than the Ki-67% score indicated by the lower CI, the proliferation score returned by your IHC staining was less than the consensus mean. Conversely, if your score was higher than the higher CI, your proliferation score is higher than the mean.

If after three runs you have consistently received reports that indicate the scores for all tissues tested are outside the CIs for the consensus mean, you should adjust your methodology to either increase or decrease the sensitivity as appropriate. However, please discuss this with your breast lead prior to making any changes.

The following FAQs are intended to help you decide what, if any action you need to take as a consequence of reviewing your quantitative results.

It is important to note that these data are also subject to the sample tissue's inherent biological variation in Ki-67, which will lead to some underlying heterogeneity in the results.

We would not suggest changing anything on the basis of one or two observations in isolation.

Before any changes in methodology are undertaken, we would always advise that they are discussed and agreed with the lead breast pathologist. And we would also caution against extrapolation of the Ki-67 results produced in breast to other clinical settings.

Frequently asked questions

I have received a set of results. All the scores are below/above their respective 95% CI. Should I change my methodology?

A set of results like this indicates that there may be a trend for your method to produce Ki-67 scores that are lower/higher than ideal.

Before taking any action, you should first consider how far away from the mean your results lie. This is where the SD measurement is useful. If all three are outside the range defined by +/- 1SD, this indicates that the current staining methodology requires alteration. If you also received assessors' comments in the qualitative assessment in-line with the quantitative ones, this increases the evidence that change is required.

Things to consider include:

- Antigen retrieval: are you using an AR of the right pH, as recommended by the primary antibody supplier? Is the duration in-line with recommendations?
- Primary antibody: if you are using a concentrate, is the dilution factor appropriate? Is the incubation duration too-short/too-long? If it is RTU, consider the incubation duration.
- Detection method: is this an up-to-date type e.g., labelled polymer/multimer?

I have received results in a number of successive assessments where the majority of the scores are consistently below/above their respective 95% CIs. Should I change my methodology?

This is another scenario that indicates that there may be a trend for your method to produce Ki-67 scores that are lower/higher than ideal.

The evidence from multiple runs should be considered as carrying more weight than that from a single assessment. Even if the scores are not outside the +/- 1SD you may wish to consider making some changes.

Is the score produced by tonsil control more or less important than the breast cancer tissue scores?

Ki-67 staining indicates that the proliferation occurring within the germinal centres of a reactive tonsil is very high – between 75 and 90%.

Thus, while the germinal centre staining is a useful 'normal' control it has inherently high biological inter-case variation that makes it less than ideal. Additionally, the fact that it has such a high proliferation index make it a control which is less useful when gauging the sensitivity of a method. In layman's terms it might be described as 'too-easy' to stain.

While we do not suggest ignoring the tonsil results, they need to be interpreted with caution and always in the context of the breast cancer control tissues.

What does a mixture of scores, some higher, some lower, and some within CI limits mean?

Some variation between individual results is to be expected. However, if you continually see such patterns where the outlying scores are not close to the 95% CIs you should examine your methodology for intrinsic sources of variation.

Our scores are very close to the CI limits but still fall outside them. Is this still a problem?

Probably not. As stated previously, when assessing your results, you should take into account the absolute scores and their distance from the cut-off, rather than their categorisation alone. It would be prudent in such instances to await the results of the next quantitative report before determining if any action is necessary.