Scientific Advisory Committee for Toxicology Laboratories

Guidelines for the Assessment of Bias and Precision in Immunoassay

Committee members:

Claire George - QuAI health

Andy King – QuAI health

Mark Parkin – Eurofins Forensic Services

Susan Grosse – Eurofins Forensic Services

Allan Traynor – Abbott Toxicology

Nicole Fernandes – Abbott Toxicology

Richard Evers – Seren Laboratory Services

Philip Kindred – Seren Laboratory Services

Samuel Hudson – Matrix Diagnostics

Mark Piper – Randox Testing Services

Paul Darkins – Randox Testing Services

Contents

1. Introduction

The following guidelines provide an overview of the recommended best practice for the assessment of performance of an immunoassay. The guidelines have been written to facilitate the consistent application and approach to measuring immunoassay performance in Toxicology laboratories for the presumptive screening of drugs of abuse in biological samples for workplace testing and medicolegal purposes. This does not preclude their use for screening of healthcare samples.

The guidelines are designed to ensure that assay performance is appropriately calculated and assessed, and that the limitations of assay performance around the cut-off are suitably presented in customer-facing documentation. This document does not replace UKAS Lab 51 and is intended to be used as an adjunct in laboratories where this applies.

For the purposes of these guidelines, immunoassay is defined as a heterogeneous or homogeneous enzyme immunoassay that is applied to the presumptive determination of drugs of abuse in biological samples. Immunoassay methods may be referred to as qualitative, or semi-quantitative. The terms "qualitative" and "semi-quantitative" are defined by the assay manufacturers and refer to the calibration model, qualitative being an assay with a single-point calibration and semi-quantitative being an assay with a multiple point calibration. The sample result is determined through the measurement of a signal generated by the reaction between the immunoassay antibody and antigen (analyte). The outcome of the testing, irrespective of the calibration model, is reported qualitatively *i.e.* negative or non-negative (positive) against a defined cut-off. Numeric values are not reported for immunoassay results.

2. Quality Control of Immunoassays

To ensure that assay performance is correctly evaluated, estimated bias and precision must be established during the validation of new assays.

Quality Control (QC) material must be homogeneous material of known origin that is metrologically traceable to an appropriate certified reference material. Where practicable, QC material should be prepared in a matrix to match that of samples being analysed. Where this is not possible, the closest available matrix should be used.

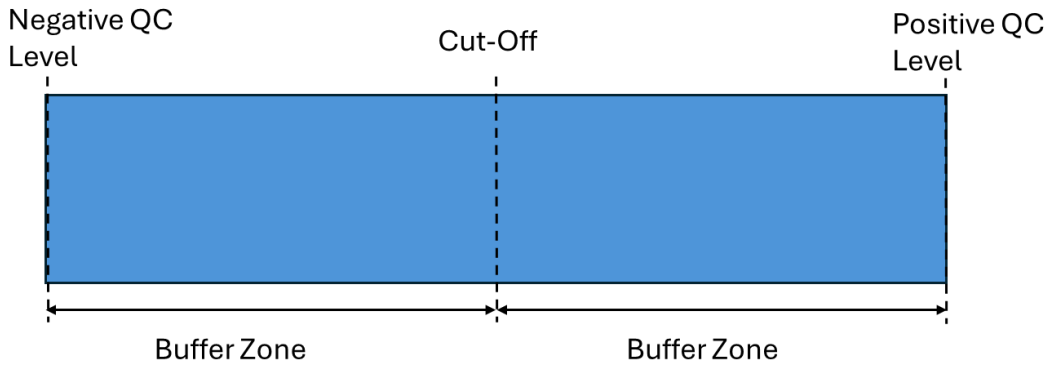
The immunoassay response at the cut-off and at each QC level must be within the linear part of the assay range.

Laboratory procedures for the assessment of on-going assay performance must ensure that there is no cross-over between the limits applied to positive and negative QCs at the ±2SD level and positive QC produces a positive result and negative QC produces a negative result.

Immunoassay performance should be monitored according to the manufacturer's recommendations provided in the applicable instructions for use (IFU). QC is typically performed at ±25% of the cut-off for urine assays and ±50% of the cut-off for oral fluid assays unless otherwise stated in the IFU. This approach means that there is a 'buffer zone' (Fig 1) where performance around the cut-off is not assessed.

The capability of an immunoassay to effectively distinguish between negative samples *i.e.* those containing no drug or drug at a level that is less than the specified cut-off, from those which contain drug at a concentration above the specified cut-off can be achieved by calculating the ratio of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). This assessment is performed using data from the Positive and Negative QCs – refer to section 5.

Fig 1: Buffer zone around the cut-off between QC levels bracketing the Cut-Off



3. Bias

An understanding of Bias at each Quality Control (QC) level provides insight into the likelihood of False Positive or False Negative outcomes.

Bias must be assessed during the validation of a new assay and at defined time intervals post validation to ensure appropriate performance is maintained.

The assessment of bias is achieved by:

- 3.1 Assessment of a minimum of 50 data points collected over a minimum of 5 days at each QC level.
- 3.2 The maximum acceptable bias is ±20% from the target QC value.
- 3.3 Where Bias exceeds ±20% this should be investigated and remedial action taken where possible.
- 3.4 Where the bias exceeds ±20% and this is not resolved by remedial action, the immunoassay may only be used for the analysis of routine samples if:
  - 3.4.1 There is no cross-over between the negative and positive QC limits at the ±2SD limits and positive QC produces a positive result and negative QC produces a negative result.
  - 3.4.2 If there is a negative bias, samples with results in the buffer zone between the negative QC and the cut-off are additionally tested via a confirmatory technique prior to reporting - where the test is performed as part of workplace drug testing services. No further action is required for healthcare samples, or if a positive bias is evident and confirmation analysis is going to be undertaken.
  - 3.4.3 The performance of the test is clearly documented in customer-facing communication (see section 7).

4. Precision

The repeatability and intermediate precision must be assessed during the validation of a new immunoassay, with the on-going reproducibility being assessed at defined time intervals (on-going assessment) post-validation to ensure appropriate performance is maintained.

The maximum acceptable CV% for repeatability must be defined within the validation protocol. The maximum acceptable CV% for intermediate precision and reproducibility is 20%.

The assessment of precision is achieved by:

4.1 **Repeatability**

4.1.1 Assessment of a minimum of **20** data points at each QC level collected under repeatability conditions

4.2 **Intermediate precision**

4.2.1 Assessment of a minimum of **50** data points collected over a minimum of **5 days** at each QC level

4.3 **Reproducibility** (on-going assessment of method performance)

4.3.1 Assessment must be performed at least annually and include a minimum of 100 data points at each QC level

5. Assessment of Assay Performance

Since immunoassays in the context of this document are qualitative, the assessment of performance is undertaken by evaluating the number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) results against a defined cut-off.

Assessment of qualitative immunoassay performance must be assessed during validation and may use the same Positive and Negative QC data set as that used to calculate Intermediate precision, if deemed appropriate. On-going monitoring of qualitative performance must be performed at least annually and may use the same Positive and Negative QC data set as that used to calculate Reproducibility, if deemed appropriate.

The assessment of qualitative immunoassay performance may include the following parameters: sensitivity, specificity, positive predictive value, negative predictive value and accuracy. These are calculated as follows:

Parameter	Calculation	Data Set
Sensitivity (%)	$\frac{TP}{TP+FN}\ x100$	Positive QC
Specificity (%)	$\frac{TN}{TN+FP}\ x100$	Negative QC
Positive Predictive Value (%)	$\frac{TP}{TP+FP}\ x100$	Authentic samples

Negative Predictive Value (%)	$\frac{TN}{TN + FN}\ x100$	Authentic samples
Accuracy (%)	$\frac{TP + TN}{Total\ number\ of\ samples\ tested}\ x100$	Positive and Negative QC

PPV values may not be available for all tests and NPV values are unlikely to be available at all. If laboratories wish to calculate PPV data it should be based off a minimum of 100 authentic samples that screen positive.

6. Measurement Uncertainty

Immunoassay screening tests present difficulties with calculation of MU due to differing responses to drugs within a group. To further explain MU laboratories should document relevant drugs that are known to cause false positive or false negative results if this information is not contained in the IFU.

Laboratory monitoring of performance should be based on the precision and bias assessments described in sections 3 & 4 above, whereas MU should be calculated and represented as per calculations in section 5.

7. Communication of Immunoassay Performance

In order to ensure transparency regarding immunoassay performance and to communicate Measurement Uncertainty, customer-facing documentation must ensure reference to the following:

7.1 Accuracy (%)

The accuracy for each immunoassay must be provided together with an explanation of what the accuracy represents *i.e.* the ability of the assay to correctly differentiate the presence or absence of a drug against a defined cut-off, and the concentration(s) used in the assessment of the accuracy.
This should be reviewed in line with a frequency stipulated in the laboratory's procedure for the assessment of Measurement Uncertainty. Customer-facing documentation must be updated to reflect any change in performance.

7.2 Assessment of Bias

Documentation must stipulate that a bias of up to ±20% is accepted and the action taken where the bias exceeds this limit *i.e.* workplace samples with results between the negative QC response and the cut-off (buffer zone) will be tested via a confirmatory technique prior to reporting. No further action will be taken on healthcare samples.

7.3 Decision rule

An explanation of the binary decision rule and simple acceptance. No guard band is applied to the results with the application of a binary decision rule *i.e.* the results are expressed as negative or non-negative (Positive). The use of a simple

acceptance rule means that the probability to be above or below the cut-off may be as high as 50% when the analyte concentration falls exactly on the cut-off.

8. Appendix 1 – Definitions

Term	Definition
Accuracy	Ability of the assay to correctly differentiate the presence or absence of a drug against a defined cut-off
Bias	The difference between the mean of a number of determinations obtained under repeatability conditions and the true or accepted concentration
Cut-off	The concentration level at which a sample is determined to be Positive for the target analyte
CV	An estimate of the relative standard deviation of a population from a statistical sample of n results divided by the mean of that sample
Decision rule	A rule that describes how measurement uncertainty is accounted for when stating conformity with a specified requirement
False Negative	A negative result obtained when the target analyte is present in a sample at a concentration equal to or greater than the cut-off
False Positive	A positive result obtained when the target analyte is not present in a sample or present but below the cut-off concentration
IFU	Instructions for use
Intermediate precision	The correlation between a series of measurements obtained from the analysis of a homogeneous material (QC) across different days by different analysts (where applicable).
Measurement Uncertainty (MU)	A measure of the dispersion of the values that could reasonably be attributed to the analyte or 'how certain are we of this result'
Negative Predictive Value (NPV)	The proportion of negative screening tests that return a confirmed negative result
Positive Predictive Value (PPV)	The proportion of positive screening tests that return a confirmed positive result
Quality Control (QC)	A homogeneous material of known origin and concentration used to assess and monitor method performance
Repeatability	The correlation between a series of measurements obtained from analysis of a homogeneous material (QC) under the same operating conditions over 1 day
Reproducibility	The correlation between a series of measurements obtained from analysis of a homogeneous material (QC) by different laboratories or across a prolonged time period
Sensitivity	Ability of the assay to correctly identify samples where the target analyte is present at a concentration at or above the cut-off
Specificity	Ability of the assay to correctly identify samples where the target analyte is not present or present below the cut-off
True Negative	A negative result obtained when the target analyte is not present or is present below the cut-off

True Positive	A positive result obtained when the target analyte is present at or above the cut-off