Guidelines for the Evaluation of PoCT Instruments that Provide Quantitative Results

Introduction

Method evaluation, validation and verification provides objective evidence that a method is fit for purpose, meaning that the particular requirements for a specific intended use are fulfilled[1].

It is not intended that every new user or purchaser of a POCT device should be performing an evaluation as described in this document. We foresee that this document will apply to evaluations of equipment which are new to the market or to evaluations that might be performed by laboratories when assessing purchase of POCT equipment to be used in their hospital or network.

This AACB Guideline originated from the AACB Point of Care Working Party and has been reviewed and endorsed by IVD Australia.

Instrument Overview

- This section should describe the intended/most appropriate use of the instrument including the anticipated clinical area where this may be used.
- Description of the analytical methodology including sample type/ volume, measuring range, specificity claims and limitations, testing interval (time to test result).
- Any patient populations for which the instrument/test should not be used.
- Limitations stated by the manufacturer
- Instrument requirements eg electricity/battery, refrigeration for consumables, infectious waste disposal
- Regulatory status (ARTG, FDA, CE)

Quality Goals

- Description of e performance goals and published analytical results for the instrument/method on specified sample populations (manufacturer package insert, independent publications)

Aim of Evaluation

- Description of the scope of the evaluation including the desirable analytical goals

Product Description

This should include the following points:

A brief description of device

- Analytical methodology

- Calibration/traceability
- Sample type
- Sample volume
- Measurement range
- Analysis time
- Quality control requirements
- Connectivity capabilities
- Precision claims
- Accuracy claims
- Any potential interfering substances
- Recommended use of device (ie on which population group)
- Maintenance requirements

In addition the limit of detection should be evaluated, especially where it is critical to clinical decision making.

General evaluation considerations

Before undertaking the evaluation the investigator should be familiar with the operation of the device as specified by the manufacturer's instructions. It is also important to check that the device is performing according to manufacturer's specifications. Any evaluation report should include evaluation of the device in the hands of the intended users eg nurses and other non-laboratory trained carers.

Ongoing analytical performance is investigated by analysing quality control samples according to manufacturer's instructions or any other recommended procedures as set by the manufacturer with assessment of precision and accuracy.

Reference should be made to CLSI documents that describe all aspects of analytical evaluations.

Any evaluation should be conducted by or at least in consultation with appropriately experienced personnel.

Evaluation of Precision

Prior to commencement of the evaluation, minimum performance specifications should be decided upon based on the intended clinical use of the test. Measurement of precision is usually expressed numerically as imprecision – standard deviation (SD) or coefficient of variation (CV). Assessment of the precision of the device should be based on both quality control material and patient samples.

- Both intra-assay imprecision (within-run) and inter-assay (between-run) imprecision should be determined. Intra assay imprecision is determined using same the lot number of consumables and the same operator over a relatively short period of time. A minimum of 20 replicates should be used. Inter-assay imprecision is determined over an extended period of time and involves multiple operators and different consumable lot numbers.

- Imprecision goals – the maximum allowable SD and/or CV(%) at clinically relevant analyte concentration should be determined.
  CLSI document EP15-A2 suggests performing one run per day with 3 replicates at each of two concentrations daily for 5 days. With patient samples only within run imprecision is usually possible, due to the potential problems with analyte stability in human material.

- An Excel spreadsheet tool for calculations is available on the AACB web site, "Doug's Pathology Utilities"[2.]

  Calculate the standard deviation for replicates and compare to the manufacturers claims.

  Imprecision, expressed as a coefficient of variation [CV%], is calculated using the formula:
  CV% = (standard deviation [SD]/mean) x 100%.

  As a general rule, the lower the imprecision, the better the reproducibility of the device.

  If the samples used in the patient comparison are tested in duplicate on the point-of-care device, then the imprecision can be estimated as follows: take the differences between the duplicates, square each difference, add all the squares, divide by twice the number of pairs of results (total number of results) and calculate the square root.

  As a minimum for PoCT intra-assay precision it is recommended to test two levels of quality control in duplicate over 5 days.

Evaluation of Linearity

Linearity is used to establish the measuring interval that can be reported for the assay under evaluation ie the upper and lower results between which all results can be reported.

When evaluating a PoCT method, the linearity quoted by the manufacturer should be confirmed by analysing a minimum of 2 replicates at 5-7 different concentrations over the claimed measuring interval (CLSI document EP06-A). Possible matrix effects (influence of substances contained or not contained in the material being used compared to the material intended to be analysed) must be excluded.

Some companies provide calibration verification kits/materials that can be used to confirm the linearity performance of their devices.

Comparison of POCT device results to other (laboratory) method

The evaluation should state details of the comparison method with to which the POCT method is to be compared. Ideally the comparison method would be a so-called reference method but this is unlikely given the difficulty of establishing such methods and instead a comparator method of known bias, such as can be judged

from external quality assessment data, should be selected. Important considerations are:

- At least 40 samples covering the clinically meaningful range should be included in the comparison. Duplicates should be run for both PoCT device and comparative methods.
- Samples should be run within a time span consistent with analyte stability. In general, the time span should not exceed 2 hours for analysis by each method. The manufacturer's recommendations on sample stability must be followed.

If any errors occur during the analysis, these results should be excluded from the final analysis and the reasons for exclusion documented. Any discrepant results must be further investigated using a third (different) laboratory method.

Analysis of Method Comparison

Before statistical evaluation is performed, a scatterplot and a difference plot should be carefully examined to identify outliers. If non-constant scatter is observed or suspected it is recommended that more than 40 comparison samples are used for analysis. The results should be used to generate the following:

i. **A Bland-Altman difference plot**
   This will calculate the average (mean) bias of the point-of-care device relative to the comparative method and the limits of agreement (limits within which 95% of the differences fall). The closer the mean bias is to zero the closer the POCT device is to the laboratory method. If a bias exists this plot will highlight if the bias is constant across the concentration range or if it is proportional to the analyte concentration.
   Any bias should be within published clinically allowable limits.

ii. **A regression analysis**
   The correlation coefficient (r) characterises the dispersion of results around the line of best fit. The closer that r is to 1, the better the fit. Some potential reasons for obtaining low r values are:
   - There is not sufficient spread of results throughout the measuring range
   - Interferences
   - Poor correlation between methods

The slope (proportional bias) indicates the angle of the line of best fit – the closer to 1 the better; it is often related to the calibration differences between methods

The Y-intercept is the point at which the line of best fit intersects the y axis (constant bias) and also may be related to calibration

A Passing Bablok regression is recommended since it allows the comparison of two analytical methods to determine systematic error or bias. It's advantage compared to least squares linear regression is that it allows for measurement error in both X and Y

variables, This means it does not assume any measurement error is normally distributed this making it more robust in case of outliers.

iii. **Error Grid analysis**

This is a useful tool in the evaluation of glucose meters. It is a clinically oriented nonparametric approach to comparing blood glucose methods, based on three assumptions: 1) glucose readings <3.9 mmol/L should be treated to raise levels, 2) glucose readings >10 mmol/L should be treated to be lowered, and 3) acceptably accurate estimates are within 20% of the reference method or when both the estimates and reference blood glucoses are <3.9 mmol/L.

The grid breaks down a scatterplot of reference glucose values versus values from the evaluated glucose meter into five regions:

1) Region A are those values within 20% of the reference method
2) Region B contains points that are outside of 20% but would not lead to inappropriate treatment
3) Region C are those points leading to unnecessary treatment
4) Region D are those points indicating a potentially dangerous failure to detect hypoglycemia or hyperglycemia
5) Region E is those points that would confuse treatment of hypoglycemia for hyperglycemia and vice-versa.

Interferences

An interference in a method is an artefactual over- or under reporting of a result due to the presence of a substance that reacts non-specifically with the measuring system. Interfering substances to be tested are selected from the manufacturer's performance claims or published reports on interfering substances which can affect the analyte of interest.

Two aliquots of identical patient sample are required to test interferences – the interfering substance being checked is added to one sample, the other sample has added to it a solution that does not contain the interfering substance. Both samples are analysed to see if there is any difference in values due to the addition of interfering substance.

- Samples should be analysed in duplicate.
- The amount of interfering substance added needs to achieve values near the maximum concentration expected in the patient population. If the interference is found at maximum concentration then lower concentrations of interfering substance should be tested to determine the level at which the interference first affects test results.
- It is recommended at least three analyte concentrations are tested for interference.

A good practice is to test (as the minimum) the following common interferences:

- Bilirubin – test by adding standard bilirubin solution
- Haemolysis – test by mechanically haemolysing part of one of the paired samples by freezing and thawing and adding back to the original at predetermined concentrations of haemoglobin.
- Lipemia – test by adding a commercial fat emulsion or by analysing a lipemic patient sample before and after ultracentrifugation
- Exogenous analytes/drugs – test by adding the analyte/drug of interest

Conclusions

When the all the various aspects of the evaluation have been performed it is worth reviewing how the results of this evaluation compare with other published evaluations of the same equipment including concordance to the manufacturer's claims. This comparison can be included in a summary a description of the analytical performance found in the evaluation and whether the results have met the accepted analytical goals for the evaluation. In addition comment should be made on how the evaluation data compares with recommendations from professional societies and whether the evaluated device can be recommended for the clinical use/population group in question.

References

1) http://www.nata.com.au/phocadownload/publications/Technical_publications/Technotes_Infopapers/technical_note_17.pdf, accessed 31 March, 2012.

2) http://www.aacb.asn.au/web/Resources/Tools/

Further Reading

1) **EP05-A2**

Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline —Second Edition

2) **EP06-A**

Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline

3) **EP09-A2IR**

Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline - Second Edition (Interim Revision

4) **EP07-A2**

Interference Testing in Clinical Chemistry; Approved Guideline - Second Edition