Guidelines for the Evaluation of PoCT Instruments

Introduction

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

Instrument Overview

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

Quality Goals

- Describe published analytical goals for instrument (manufacturer package insert, independent publications)

Aim of Evaluation

- Describe scope of evaluation including desirable analytical goals

Product Description

Brief description of:

- description of device
- methodology
- calibration/traceability
- sample type
- sample volume
- Measurement range
- Analysis time
- Quality Control requirements
- Connectivity capabilities
- Manufacturer precision claims
- Manufacturer accuracy claims
- Interfering Substances that operator should be aware of
• Recommended use of device (ie population group)
• Maintenance requirements

**Evaluation Preparation**

Before undertaking evaluation the investigator should be familiar with operation of device as specified by manufacturer’s instructions.

Before commencing evaluation it is important that the device is performing according to manufacturer’s specifications. Device optimal performance is investigated by running quality control according to manufacturer’s instructions or any other recommended procedures as set by the manufacturer.

The evaluation report should include an evaluation of the device in the hands of the intended users.

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

**Precision**

Prior to commencement of 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). Results obtained from precision studies need to be assessed to ensure they meet clinical specifications set prior to evaluation.

Precision profile of the device should include both quality control material and patient samples as much as possible.

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

• Imprecision goals – maximum allowable SD and/or CV(%) at clinically relevant analyte concentration should be determined. CLSI document EP15-A2 suggests running 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 nature of the fresh human material.

Excel spreadsheet tool is available on the AACB web site, “Doug’s Pathology Utilities”...

Calculate standard deviation for replicates and compare to manufacturers claims

Imprecision, expressed as a coefficient of variation [CV%], is calculated using the formula:

\[ CV\% = \frac{\text{standard deviation [SD]}}{\text{mean}} \times 100\% \]

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 up all the squares, divide by twice the number of pairs of results (total number of results) and take the square root.

It is recommended that imprecision studies are also performed using quality control material.

**Linearity**

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

When evaluating a PoCT method linearity quoted by manufacturer should be confirmed by running a minimum of 2 replicates at 5-7 concentrations over the claimed measuring interval (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.

**Method Comparison**

If the device is compared to an approved reference method the difference between the PoCT device and the comparative method measures the trueness of the PoCT device. If the comparative method is not a reference method (e.g. a standard laboratory method) trueness cannot be claimed. Bias reported will be “bias to comparative method”. The appropriate way to determine which method delivers the correct result is to compare to a reference method.

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

While running comparisons all errors should be documented. Data attached to documented errors should not be included in final analysis. Any cause that requires rejection of data should be documented. Any discrepant results must be further investigated using a third (different) laboratory method.
Analysis of Method Comparison

Plot a scattergram of results. If non-constant scatter is observed or suspected it is recommended that more than 40 comparison samples are used for analysis.

i. **Construct a Bland-Altman difference plot.**
   This plot 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.
   Bias should be within published clinically allowable limits (reference [QAP ALs](#)).

ii. **Perform a regression analysis.**
   Correlation coefficient \( r \) characterises the dispersion of results around the line of best fit. The closer to one \( r \) is the better the fit. Potential reasons for obtaining low \( r \) values:
   - Inadequate range of values – good indicator if inadequate range of samples used for evaluation
   - Interferences
   - Poor correlation between methods

   Slope (proportional bias) indicates the lean of the line of best fit – the closer to one the better; often related to calibration differences between methods

   Y-intercept is the point at which the line of best fit intersects the y axis (constant bias) – the closer to zero the better; may be related to calibration

   A Passing Bablok regression is recommended since it allows you to compare two analytical methods to determine systematic error or bias. Advantage of using a Passing Bablok regression over least squares linear regression is that it allows measurement error in both X and Y variables, This means is does not assume measurement error is normally distributed making it robust against outliers.

iii. **Error Grid analysis**
   A useful tool in evaluations of glucose meters is an error grid analysis. The error grid is useful to quantify the clinical accuracy of a glucose meter compared to a reference value. 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 raised, 2) glucose readings >10 mmol/L should be lowered, and 3) acceptably accurate estimates are within 20% of the reference method or when both the estimates and reference BG are <3.9 mmol/L.
The grid breaks down a scatterplot of a reference glucose method and an 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.

4. Interference

Interference 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.
Substances to be tested are selected from the manufacturer’s performance claims or published reports on interfering substances which affect analyte of interest.
Two aliquots of identical patient sample is required to test interference – interfering substance being checked is added to one sample, the other sample has added to it a solution that does not contain interfering substance. Both samples are analysed to see if there is any difference in values due to addition of interfering substance.

- Samples should be analysed in duplicate.
- Amount of interfering substance added needs to achieve values near expected maximum concentration expected in patient population. If 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

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 analyte/drug of interest

Conclusions
Discuss how results of this evaluation compares with other published evaluations of the same equipment including concordance to manufacturer’s claims.

Describe how analytical performance compares with recommendations from professional societies

Recommend the clinical use/population group (if any) that the instrument is suitable.

References.