Evaluating Assay Precision

Douglas Chesher
Department of Clinical Biochemistry, Pacific Laboratory Medicine Services, Royal North Shore Hospital,
St Leonards NSW 2065, Australia.
For correspondence: Dr Douglas Chesher e-mail: dougc@med.usyd.edu.au

Summary

- When evaluating the precision of a method it is necessary to assess the repeatability (within-run) and the total or within-laboratory precision.
- It is insufficient to assess repeatability in a single run.
- Clinical and Laboratory Standards Institute (CLSI) document EP05-A2 describes the protocols for determining the precision of a method.[1] The precision of a method should be tested at at-least two levels; each run in duplicate, with two runs per day over 20 days.
  CLSI document EP15-A2 describes the protocols that should be undertaken by the user to verify precision claims by a manufacturer. Precision claims by a manufacturer should be tested at at-least two levels, by running three replicates over five days.[2]
- A spreadsheet for assisting with the calculations described in this article is available from the AACB web-site.[3]

Introduction

Part of the process of verifying or validating a method to confirm that it is suitable for use is an assessment of precision. By this we mean the closeness of agreement between independent results of measurements obtained under stipulated conditions; it is solely related to the random error of measurements and has no relation to trueness/accuracy.[4]

There is some variation in the terminology used but for the purposes of this discussion, repeatability, also known as within-run precision, is defined as the closeness of agreement between results of successive measurements obtained under identical conditions. Reproducibility is at the other extreme and refers to the closeness of agreement between results of successive measurements obtained under changed conditions (time, operators, calibrators, reagents, and laboratory). For the purposes of this discussion reproducibility will not be considered, as it involves multiple laboratories. Instead total precision within a laboratory (within-laboratory precision) will be assessed.

While the term precision relates to the concept of variation around a central value, imprecision is actually what is measured. For a normal distribution the measure of imprecision is the standard deviation (SD). Alternatively one can use the variance, which is simply the square of the SD.

It is generally assumed in the laboratory that the variation associated with repeated analysis will follow a normal distribution, also known as the Laplace-Gaussian or Gaussian distribution. If this is true then using the principle of analysis of variance components:

$$\sigma^2_{Total} = \sigma^2_{Within-run} + \sigma^2_{Between-run}$$

where σ is the SD and σ^2 the variance.

Note, some authors refer to total variation as just the between-run component instead of combined between-run and within-run shown above. Care must be taken in knowing which term is being referred to. CLSI now uses the term within-laboratory precision to denote the total precision within the same facility using the same equipment[1] and this term will be used for this concept throughout this paper.

In a production like process, such as measuring an analyte in some matrix, the mean (μ) and SD (σ) are not known, and can only be estimated. The estimates of μ and σ are usually denoted by $\bar{x}$ and s respectively. For n measurements we have:

$$\bar{x} = \frac{\sum x}{n} \quad (1)$$

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} \quad (2)$$

The coefficient of variation (CV) is defined as:

$$CV = \frac{s}{\bar{x}} \times 100\% \quad (3)$$

Estimating Precision

When evaluating the precision of an assay, the trivial approach for estimating repeatability for any given level is to perform 20 replicate analyses in a single run on a single day. Similarly the within-laboratory precision is estimated by measuring a sample 20 times over multiple days. Unfortunately this approach is insufficient, as it tends to under-estimate repeatability, as the operating conditions in effect at the time may not reflect usual operating parameters.[2]

CLSI recommends an alternative approach that is described in documents EP05-A2[1] and EP15-A2.[2] The two documents are intended for different purposes. EP05-A2 should be used to validate a method against user requirements, and is generally used by reagent and instrument suppliers to demonstrate the precision of their methods. In contrast, EP15-A2 is intended to verify that a laboratory's performance is consistent with claims made by the manufacturer. Thus a laboratory may choose to use an EP15-A2 based assessment if it is verifying a method on an automated platform using the manufacturer's reagents. However, for a method developed in-house a higher level of proof is required to validate the method, in which case EP05-A2 would be the appropriate guideline to use.

Various materials may be used to complete the assessment with either protocol. These include pooled patient samples, quality control material, or commercial standard material with known values. When using quality control samples, these should be different to those used to ensure the instrument is in control at the time of the assessment.

As the period of assessment is quite short, the total SD or within-laboratory SD derived from these experiments should not generally be used to define acceptability limits for internal quality control. For this, longer-term assessment is required.

Assessment Protocols

The EP05-A2 protocol recommends that:

- The assessment is performed on at least two levels, as precision can differ over the analytical range of an assay.
- Each level is run in duplicate, with two runs per day over 20 days, and each run separated by a minimum of two hours.
- There should be at least one quality control (QC) sample in each run. If QC material is being used for the precision assessment, it should be different to that used to control the assay.
- The order of analysis of test materials and QC for each run or day should be changed.
- To simulate actual operation, include at least ten patient samples in each run.

The EP15-A2 protocol is similar except that the experiment is undertaken with three replicates over five days for at least two levels. The reader is referred to the CLSI documents for details.[1,2]

When undertaking the assessment the data must be assessed for outliers, which are considered to be present if the absolute difference between replicates exceeds 5.5 times the SD determined in the preliminary precision test. If an outlier is found the pair should be rejected and the cause investigated and resolved before repeating the run. The figure of 5.5 is derived from the upper 99.9% value of the normalised range for differences between two populations.

Estimation of Repeatability and Within-Laboratory Precision

The following example relates to the verification of performance of calcium according to EP15-A2 using a five day protocol. For the purposes of this example the results of only a single level are shown (Table 1).

Table 1. Calcium results for level 1 (mmol/L).

Run / Day	Replicate 1	Replicate 2	Replicate 3
1	2.015	2.013	1.963
2	2.019	2.002	1.979
3	2.025	1.959	2.000
4	1.972	1.95	1.973
5	1.981	1.956	1.957

Repeatability

Repeatability is estimated using the equation below.

$$s_r = \sqrt{\frac{\sum_{d=1}^{D}\sum_{r=1}^{n}(x_{dr} - \bar{x}_d)^2}{D(n-1)}} \quad (4)$$

where:

D = total number of days.

n = total number of replicates per day.

x_{dr} = result for replicate r on day d.

$\bar{x}_d$ = average of all replicates on day d.

The first step is to calculate the mean of the replicates for each day, then for each result subtract the mean for that day

and square the resultant value. For example, on day 1 the average of the three values is (2.015 + 2.013 + 1.963)/3 = 1.997. The first replicate on day 1 is 2.015, so we calculate $(2.015 - 1.997)^2 = 0.00032$.

Table 2 shows the results of each of these calculations.

Table 2.

Mean for day	(R1 - Mean)²	(R2 - Mean)²	(R3 - Mean)²
1.997	0.00032	0.00026	0.00116
2.000	0.00036	0.00000	0.00044
1.995	0.00092	0.00127	0.00003
1.965	0.00005	0.00022	0.00006
1.965	0.00027	0.00008	0.00006

The sum of the squared differences is 0.0055, and we know D=5 and n=3. Thus:

$$s_r = \sqrt{\frac{0.0055}{5 \times (3-1)}} = 0.023 \text{ mmol/L}$$

The next step is to calculate the variance for the daily means (s_b^2) using the equation.

$$s_b^2 = \frac{\sum_{d=1}^{D} \left(\bar{x}_d - \bar{\bar{x}}\right)^2}{D-1} \quad (5)$$

where:
D = total number of days.
$\bar{x}_d$ = average of all replicates on day d.
$\bar{\bar{x}}$ = average of all results.

Using the values from our example the mean of all the results is 1.984 mmol/L. On day 1 the mean of the three replicates was 1.997 (see Table 2), so the square of the difference from the overall mean is $(1.997 - 1.984)^2 = 0.000162$.

Table 3 shows the results of the same calculation for the remaining days.

Table 3.

Mean for day	$(\bar{x} - \bar{\bar{x}})^2$
1.997	0.000162
2.000	0.000247
1.995	0.000108
1.965	0.000371
1.965	0.000384

Summing the square of the differences gives a total of 0.00127. Thus the variance of the daily means is:

$$s_b^2 = \frac{0.00127}{4} = 0.000318$$

Within-Laboratory Precision

Finally, we can calculate the total or within-laboratory SD (s_l) using the equation:

$$s_l = \sqrt{\frac{n-1}{n} \bullet s_r^2 + s_b^2} \quad (6)$$

where n is the number of replicates per day.

Again, using the values from the example:

$$s_l = \sqrt{\frac{3-1}{3} \bullet 0.000529 + 0.000318} = 0.026 \text{ mmol/L}$$

Evaluation of Results

As alluded to above, EP15-A2 is generally used to verify that a method is performing as is claimed by the manufacturer. Therefore the imprecision estimates calculated above must be compared to the manufacturer's claim. If the repeatability and within-laboratory SD are less than that indicated by the manufacturer, then the user has demonstrated precision consistent with the claim and no further calculations are required. However, if the values achieved are greater than those reported by the manufacturer, a statistical test needs to be performed to determine whether this difference is statistically significant.

Repeatability Verification Value

In order to compare the estimated repeatability to a claimed value we can calculate the critical or verification value using the equation:

$$\text{Verification Value} = \frac{\sigma_r \bullet \sqrt{C}}{\sqrt{v}} \quad (7)$$

where:
σ_r is the claimed repeatability.
C is the 1-α/q percentage point of the Chi-square distribution.
α is the false rejection rate and q is the number of levels tested
v is the degrees of freedom and in this instance is equal to D•(n-1).

Using the example data and assuming the claimed repeatability is an improbable CV of 1.1%, the claimed SD is therefore 1.984/0.01 = 0.022 mmol/L which is less than the estimated repeatability of 0.023. In order to calculate the verification value we must first calculate the repeatability degrees of freedom, which is equal to D•(n-1). For the EP15-A2 protocol of three replicates over five days we get v = 10.

To calculate C, α is traditionally taken as 5% and q is equal to 2. Thus we need to find the 97.5% percentage point of the Chi-

square distribution with 10 degrees of freedom. This can be looked up using statistical tables or by use of the spreadsheet function CHIINV(α/q,v) = CHIINV(0.05/2,10) = 20.48.[3]

Therefore the verification value is:

$$\text{Verification value} = \frac{0.022 \bullet \sqrt{20.48}}{\sqrt{10}} = 0.031$$

As the estimated repeatability is less than or equal to the verification value the data are consistent with the manufacturer's claim.

Within-Laboratory Precision Verification Value

An analogous method is used to compare the estimated within-laboratory SD to the manufacturer's claim with the verification value being calculated according to equation 8.

$$\text{Verification value} = \frac{\sigma_l \bullet \sqrt{C}}{\sqrt{T}} \quad (8)$$

where σ_l is the claimed within-laboratory or total SD and T is the effective degrees of freedom for the within-laboratory precision estimate. T is best calculated in a spreadsheet and is given by:

$$T = \frac{((n-1) \bullet s_r^2 + (n \bullet s_b^2))^2}{(\frac{n-1}{D}) \bullet s_r^4 + (\frac{n^2 \bullet (s_b^2)^2}{D-1})} \quad (9)$$

For our example, assume a manufacturer's claim of 1.2%, which corresponds to a SD of 0.024. Therefore:

$$\text{Verification value} = \frac{0.024 \bullet \sqrt{20.48}}{\sqrt{12.1}} = 0.031$$

Again, as the within-laboratory SD is less than or equal to the verification value the data are consistent with the manufacturer's claim.

Acknowledgements

Special thanks to Amanda Caswell for her careful review of the manuscript.

Competing Interests: None declared.

References

1. Clinical and Laboratory Standards Institute. Evaluation of precision performance of quantitative measurement methods. CLSI document EP05-A2. Wayne, PA, USA: CLSI; 2004.
2. Clinical and Laboratory Standards Institute. User verification of performance for precision and trueness; approved guideline – second edition. CLSI document EP15-A2. Wayne, PA, USA: CLSI; 2005.
3. Australasian Association of Clinical Biochemists Website. Resources – Tools. www.aacb.asn.au (Accessed 7 April 2008)
4. Linnet K, Boyd JC. Selection and analytical evaluation of methods with statistical techniques. In Burtis CA, Ashwood ER, Bruns DE. Tietz Textbook of Clinical Chemistry and Molecular Diagnostics 4th ed. Elsevier Saunders, St Louis. 2006. p357.