Calculations and charts in Equalis' result reports

Equalis, P016-eng, version 1.0

Table of contents

Introduction

Equalis organizes external quality assessment (EQA) schemes in the entire field of laboratory medicine. Many of the EQA schemes handle quantitative results, such as measuring the concentration of a specific substance in plasma or blood. A round in such an EQA scheme usually involves dividing a sample material into subsamples that are sent to several participants who analyze the material and report back the analysis results. Equalis compiles the results and sends result reports to the participants with information on how well their results match those of other participants and the expected result. Sometimes a range is specified within which the result should lie for the analysis quality to be considered optimal.

In the result reports, descriptive statistics are often used. By assuming that the reported analysis results in an EQA round are a sample from a normally distributed population (of an infinite number of possible results), the analysis results (the sample) can be used to calculate the mean value and spread. The accuracy of the estimation of these statistical measures increases as the sample size grows, i.e. when more analysis results are included.

This document describes the calculations performed by Equalis in connection with the result compilations, as well as the different types of charts used in Equalis' result reports to graphically illustrate the results. Additionally, a brief description of some basic statistical concepts is provided. The document is primarily aimed at those working with quality assessment of laboratory analyses and interested in learning more about how Equalis compiles results within the EQA schemes with quantitative results.

Equalis result compilations

Calculation of mean value and spread

In Equalis' result compilations, the mean value and the spread of participants' results are often calculated. Each participant receives a measure of how much they deviate from the mean value and information about the spread, i.e. how all participants' (anonymous) results are distributed. The spread is expressed as a standard deviation (SD) and coefficient of variation (CV). Before calculations are made, the results are usually divided into different subgroups depending on the measurement method used by the participants, making it possible to detect systematic differences between different methods. The mean value and the spread are calculated separately for each group and for all results together. In some cases, Equalis may advise participants to compare their results only with those from their own output group, especially when the results in the separate groups are normally distributed, but all results together are not (e.g. due to method differences). If the number of participants in a group is less than four, the spread is not calculated as the sample size is considered too small to provide reliable results.

Sometimes, one or more participants report results that deviate extremely from the others, possibly due to sample mix-up or temporary equipment failure. These significantly deviating results are called outliers. To obtain a more accurate calculation of the mean value and the spread, minimizing the influence of any outliers, Equalis uses the robust calculation method Algorithm A, recommended in an international standard (ISO 13528:2022). The method involves several iterations where the median value is used to calculate threshold values beyond which results are replaced by the threshold values. Preliminary mean value and SD are then calculated to determine new threshold values, and this process is repeated until the threshold values no longer change.

Algorithm A

1. **Identify the median value of all original results.**
2. **Calculate the absolute deviation from the median for each original result.**
3. **Calculate the median deviation (the median of the deviations from step 2).**
4. **Calculate preliminary SD (SD*) as: SD* = median deviation × 1.483.**
5. **Calculate threshold values: median - 1.5 × SD* and median + 1.5 × SD*.**
6. **Replace results outside the thresholds with the nearest threshold value.**
7. **Calculate preliminary mean (m*) and preliminary SD (SD*) for the new results in the usual way.**
8. **Calculate new threshold values: m* + 1.5 × 1.134 × SD* and m* - 1.5 × 1.134 × SD*.**
9. **Replace original results outside the thresholds with the threshold values from step 8.**
10. **Calculate new preliminary mean (m*) and SD*.**
11. **Repeat steps 8–10 until the robust mean and SD do not change from one iteration to the next (to the third significant digit). The final mean is the mean (m*) from the last iteration, and the final SD is 1.134 × SD* from the last iteration.**

Comparison to expected result

Each participant's result is compared to a value appointed by Equalis to a property of the sample material. This value is referred to as the "Expected result". The expected result should be as close to “the true value" as possible given the circumstances.

The expected result for **ratio scale** results is based on one of the following methods:

Method	Description
Method 1	A calculated value based on a known addition of the component to the sample material.
Method 2	A reference value based on measurements at one or more independent reference laboratories. This is considered the best way to determine an assigned value, but is often costly and limited to those components where reference methods are available. Therefore, reference values are only occasionally used.
Method 3	A value from calibration against the reference value in a CRM (certified reference material).
Method 4	A value based on measurements at a number of expert laboratories selected by Equalis.
Method 5	A value based on the mean value of group means.
Method 6	Participant’s results are compared to their group’s mean value.
Method 7	An overall mean value based on all results (consensus mean).

If any of methods 1–5 is used, it is called an assigned value. Assigned value is used as a collective term for an expected result that is not based on the average of the participants' results.

For measurements that give results on an **ordinal/nominal scale**, the expected result can be based on one of the following:

Method	Description
Method 1	By expert judgment/analysis in expert laboratories.
Method 2	By use of reference materials.
Method 3	From knowledge of the origin or preparation of the sample material.
Method 4	Using the mean or median from participant results (the median is only appropriate for the ordinal scale).

Other methods that may prove to provide reliable results may also be used.

Uncertainty of expected result

Uncertainty is a general concept that encompasses various sources of error and variability in measurements. It characterizes the collection of values within which the true value lies with a certain degree of certainty. It can be described, for example, using a standard deviation or a confidence interval. For results on the ordinal/nominal scale, no uncertainty is specified.

Below is how Equalis calculates the uncertainty in an expected result.

Method	Description
Method 1 Calculated from known additive	Described and calculated in the supplier's documentation according to general principles for standard uncertainty (uncertainty in gravimetric and volumetric measurements and the purity of any materials used in the formulation).
Method 2 Reference lab	Described and calculated in the supplier's documentation according to general principles for standard uncertainty (based on the uncertainty of the method and other uncertainties during the measurement series)
Method 3 Value from calibration against the reference value in a CRM (certified reference material)	$$u_{char} = \sqrt{u_{crm}^2 + u_{\bar{d}}^2}$$,where u_{crm} is the specified standard uncertainty for CRM and $u_{\bar{d}}$ is the standard uncertainty for $\bar{d}$ $(i.e. \frac{sd}{\sqrt{p}})$.
Method 4 Mean value from selected expert laboratories	95% confidence interval for the mean, $$mean \pm 1{,}96 * \frac{sd}{\sqrt{n}}$$ sd = sd calculated on input values n = number of groups
Method 5 Mean of group means	Same as Method 4.
Method 6 Overall mean	Uncertainty is calculated by: $2 * 1{,}25 * \frac{sd}{\sqrt{p}}$ Which is an expanded uncertainty with coverage factor k=2 (approx. 95% confidence interval). sd = Robust sd from algorithm A, p = number of results
Method 7 Ouput group mean	Same as Method 6

Quality assessment of individual results

The standard procedure for external quality assessment involves comparing each participant's result with the expected result set by Equalis for the components to be measured or examined in the sample material.

The difference (deviation) between the participant's response and the expected result is then used in the assessment of the result. For assessment of the quality of individual results, pre-set criteria are used in the form of maximum permissible deviation (acceptance limits). Results within the acceptance limit are considered to have satisfactory quality.

Components on a ratio scale

Either an Equalis quality goal is specified for the component and/or the robust standard deviation of the round is used as the criterion. Participants themselves may specify which criterion to use.

<u>Equalis' quality goals</u>

Equalis' quality goals are specified in U040 and are usually determined by the Equalis' advisory group in the field. U040 is available only in Swedish on Equalis website: https://www.equalis.se/sv/deltagare/deltagarinformation/bedomning-av-resultat/

The aim of Equalis is that quality goals should be introduced into all quality assessment schemes where applicable, and that they are established in dialogue with Equalis' advisory groups and the rest of the profession.

Examples of quality goals (*QG*):

- Standard is a quality goal in % from expected result (e.g. P—Albumin +/- 5%).
- An absolute deviation based QG is used for some components (e.g. P—pH +/- 0.02)
- QG expressed in SD (z-score), e.g. for B—Leukocytes in transfusion medicine (+/- 3 SD)
- Composite limit for B—HbA1c (IFCC) +/- (1.5 + 1.65 x SD), i.e. one absolute and one based on SD.

<u>Deviation in SD (z-score)</u>

In general, the acceptance limit +/- 3 SD can be used as a replacement for +/- QG. The acceptance limit +/- 2 SD is also stated in reports from Equalis. Rules for how these limits should be used are determined by participants themselves, but generally +/- 2 SD is interpreted as a warning limit and +/- 3 SD as an action limit.

It should be noted that high overall precision in the results makes z-score based QG narrow and vice versa.

Components on ordinal or nominal scale

In most cases, the expected result for the component is pre-determined. When an expected result is available, the criterion "According to expected result" is used, i.e. if the participant's results are exactly the same as the expected result. If this is not the case, the result deviates. In some EQA schemes, the criterion "Acceptable result" is also used, which indicates that the result deviates from the expected result, but the deviation is of minor nature.

Equalis also has a few other quality goals set for certain components. Examples of quality goals in general are:

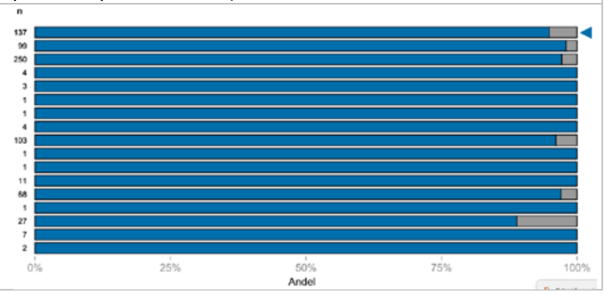
- Common ones are "No deviation from expected result" (e.g. DNA genotype).
- No false negative results (e.g. P—HIV 1+2 (ak+ag)).
- ± 1 step from the median value (e.g. Quantification of ABO antibodies).

Charts in Equalis' result reports

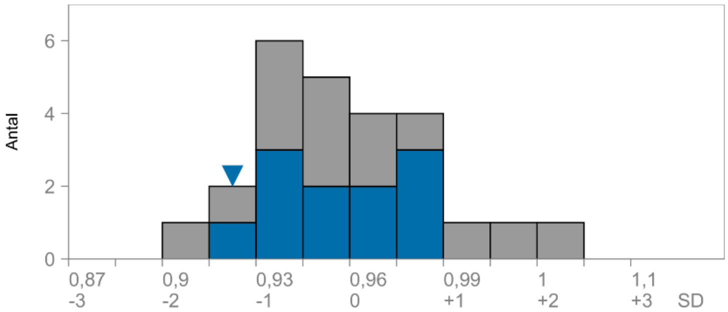
Equalis' results are presented to participants in the form of various reports, the content of which is adapted to the different EQA schemes. Below are the most commonly used charts explained.

Bar chart

The bar chart shows numbers per category, i.e. bars where the height is proportional to the number of results within each category. A variant is also used to show the proportion of different answer options per category.

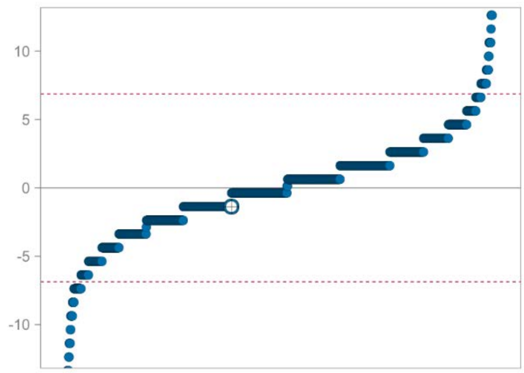
Bar chart showing number per category	*Bar chart showing proportion (e.g. positive/negative per sample material).*
	

Histogram



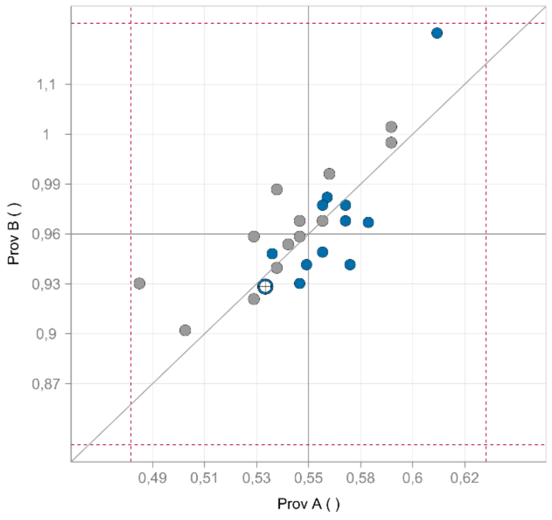
The histogram is a bar chart that shows the distribution of results, i.e. the number of results in different concentration intervals. The results are illustrated with bars where the height is proportional to the number of results within each interval.

Deviation charts



This type of chart illustrates how much the participants' results deviate from a given value. All participants' deviations are shown in ascending order along the x-axis. The y-axis shows the magnitude of the deviation. In the chart, quality goals can be plotted as lines at predetermined acceptance limits within which the analysis quality is considered optimal.

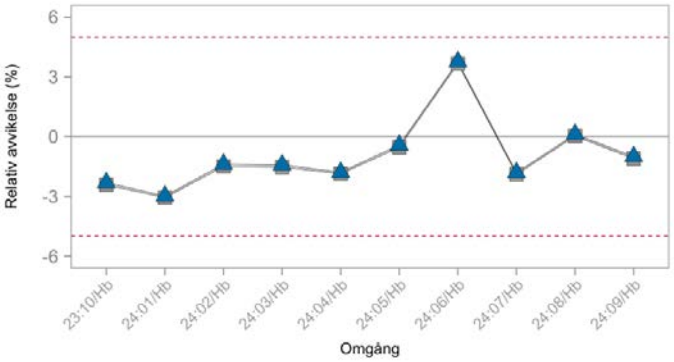
Youden plot



A Youden plot can be used to display analysis results for two different samples in the same chart. The analysis results from the sample materials A and B (with different concentrations) are plotted against each other, with the analysis results for sample A on one axis and for sample B on the other axis. Each participating laboratory is represented in the chart by a point whose coordinates are the analysis results for samples A and B.

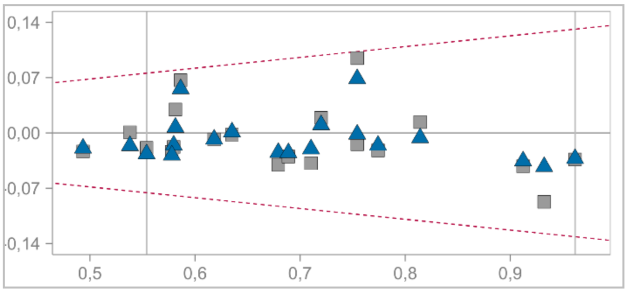
Together, the results of all participants form a point swarm, the appearance of which can be helpful in detecting any correlation between deviating analysis results and concentration level. If quality goals are available, these are plotted for both samples as red lines.

Shewhart chart



A Shewhart chart is used to track the analysis quality of an individual participant over time. The chart has an x-axis for time (e.g. date or round number for a EQA shipment) and a y-axis for the deviation between individual analysis results and the expected value. Quality goals can also be plotted in the Shewhart chart, shown as red dotted lines.

Level/concentration chart



Shows how the participant's results have deviated from a given value in relation to the concentration of the component (level) in the control material. Deviations from the most recent EQA rounds are shown in the diagram and the current round is marked with a gray line.

Deviations that vary at different levels (concentrations) may indicate that the deviations are concentration dependent.

Basic statistical concepts

Distribution

The distribution of a set of values describes how many times each value occurs. When compiling analysis results from an EQA round, the distribution of the analysis results is often described with a histogram.

Mean value

The arithmetic mean value is used to estimate the location, or "typical" value, of a distribution. It is calculated by summing all the individual observation values and dividing the sum by the number of values.

Mathematically, this is expressed as: $m = \frac{\sum x}{n}$

where m is the mean value, n is the number of values, and $\sum$x is the sum of all (n) values.

Median value

The median is another way of estimating the location of a distribution. It is a measure of centrality, around which the other values are distributed with an equal number of values on either side. The median is particularly well suited to indicate the position of skewed distributions. For perfectly symmetrical distributions, the median is the same as the mean.

Standard deviation

The standard deviation (SD) provides information about the spread of the results, i.e. the width of the distribution. It is a measure of how much the results deviate from the mean value.

Mathematically, this is expressed as: $SD = \sqrt{\frac{\sum (x-m)^2}{n-1}}$

where SD is the standard deviation, $\sum(x - m)^2$ is the sum of all squared deviations from the mean (m) and n is the number of deviations.

Coefficient of variation

The coefficient of variation (CV) is a normalized way of describing the spread of results, expressed as the standard deviation (SD) in relation to the mean (m). The coefficient of variation is often expressed as a percentage.

It is used to compare the spread at different levels/concentrations (since SD varies depending on the size of the mean).

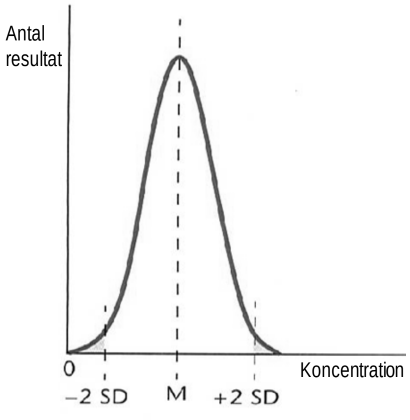
Mathematically, this is expressed as: $CV = \frac{SD}{m} \cdot 100$

Normal distribution

The normal distribution is a statistical distribution that is symmetrical around its mean. The distribution is characterized by approximately 68% of the values being within +/- 1 SD, approximately 95% within +/- 2 SD and approximately 99.7% within +/- 3 SD of the mean.

The distribution function is bell-shaped and can be calculated mathematically with the mean (M) and the standard deviation (SD).

Participant measurement data on the ratio scale are often assumed to be sampled from some (unknown) normal distribution.



Confidence interval (interval estimation)

Interval estimation is based on a percentage that defines how likely it is that the result is within an interval from a point estimate (e.g. mean value). This percentage is called the confidence level. If we have a confidence level of 95%, 5% of the cases will fall outside the interval.

References

1. ISO 13528:2022 Statistical methods for use in proficiency testing by interlaboratory comparison, Annex C.