Common Misconceptions in Medical Laboratory Quality Control

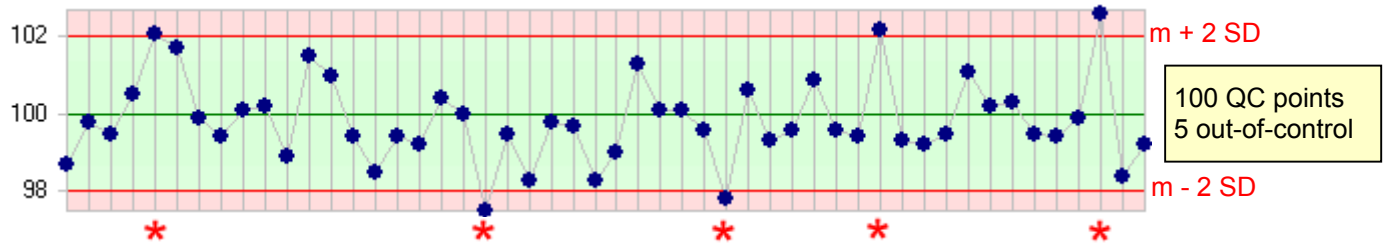
Philippe Marquis, MD
Service de biochimie, Centre hospitalier régional
Metz - France
www.multiqc.com

1. The 2 SD control Interval

Historical background

Eighty years ago, Walter Shewhart was an engineer in the electrical industry and created the control chart. He experimentally discovered that a control interval ranging between [mean ± 3 SD] was the best balance between a too high number of false rejections and a good sensitivity to notify out-of-control situations.

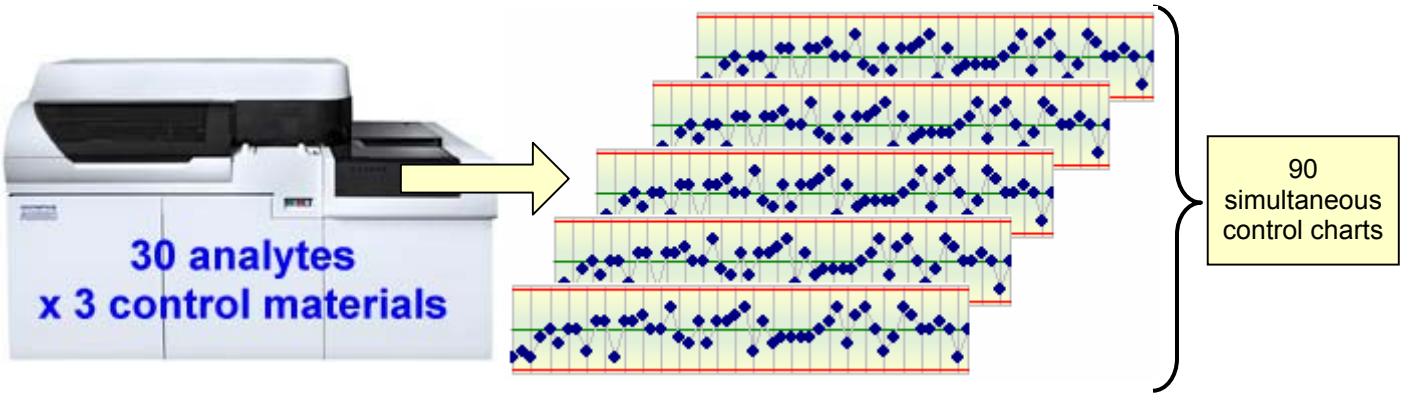
Control charts were introduced in clinical chemistry thirty years later. At this time some authors thought that a narrower control interval [mean ± 2 SD] would be a better choice because of an increased sensitivity and despite a theoretical frequency of false rejections ranging to 1/22. This narrowed control interval was maybe efficient in manual laboratory work or when automatic analysers counted only a few channels.



A 2 SD control chart has an average of 1 out of 22 points (4.5%) that necessarily falls out of the control interval

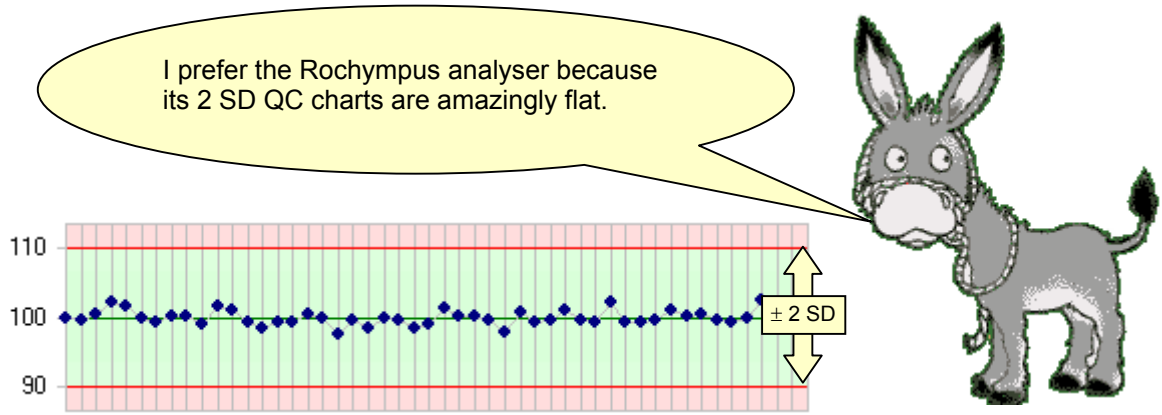
Presently

The « 2 SD » rule is completely out of date. Controlling a 30-channel analyser with 3 levels of control materials produces 90 simultaneous QC charts. A rate of 1/22 false rejections would lead to an average of 4 alarms per control run, actually a bit less because of a possible correlation between control levels. Each one of these out-of-control situations would require an investigation. So starting routine work would become impossible of very delayed. Renaming rejections as warnings does not change the root of the matter.



The « 2 SD » rule would produce an average of 4 alarms per control run of a multi-channel analyser. Working becomes impossible or very delayed.

However, several famous makers of clinical chemistry analysers (Bayer, Stago …) are delivering QC software where the entry field « Control interval » is inappropriately entitled « 2 SD interval ». Some technicians look very happy with the « 2 SD » rule. More amazing, someone in a well-known Internet forum has recently recommended an analyser because of its QC which never goes out of a 1 SD interval. Obviously the claimed SD is not the true estimated SD, but a much greater arbitrarily assigned SD. These technicians are actually performing 3 SD (or more) shewhart's charts without being conscious of it.



2. The uncertainty of control intervals

Every clinical chemist knows how to calculate the confidence interval of a mean but very few are able to do the same for a standard deviation. Nevertheless it is an easy task using the following formula :

$$SD\sqrt{\frac{n-1}{\chi^2_R}} < \sigma < SD\sqrt{\frac{n-1}{\chi^2_L}}$$

σ : true standard deviation

SD : estimated standard deviation

n : sample size

For a 95 % confidence interval :

χ^2_R : χ^2 value for n-1 degrees of freedom and P = 0.975

χ^2_L : χ^2 value for n-1 degrees of freedom and P = 0.025

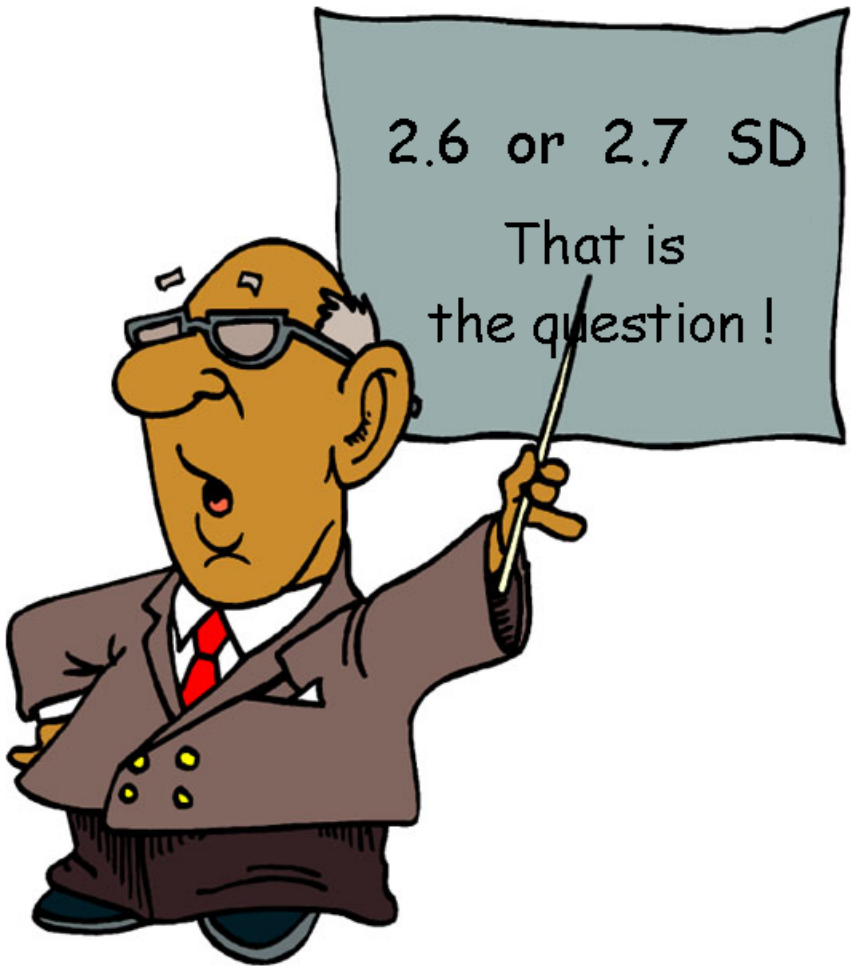
Let us apply the formula to the 3σ control chart for different sizes of reference pool often met in medical laboratory QC.

Size of the reference pool	95% confidence interval of the 3σ estimate
10	2.1 to 5.5 SD
20	2.3 to 4.4 SD
30	2.4 to 4.0 SD
50	2.5 to 3.7 SD
100	2.6 to 3.5 SD

The previous table shows that getting a reliable estimate of the 3σ control interval requires a much bigger reference pool than it is commonly advocated in clinical chemistry. Let us remember that Shewhart in 1931 recommended a reference pool made of 400 pieces (100 samples of 4 pieces). On the other hand, medical laboratory QC textbooks are generally recommending a reference pool of 30 samples.

It is important to bear in mind that under these conditions the true 3 σ control interval is something unknown between 2.4 SD and 4.0 SD. The uncertainty would be still greater if we also took into account the confidencce interval of the mean.

So do not be taken in by a famous software that claims to optimize the QC rules to meet given specifications. One can only smile when faced with such a precise computation of power functions which is based on so imprecisely estimated standard deviations. It is not wrong but who would pay for a micrometer to measure the thickness of a down-filled sleeping bag ?

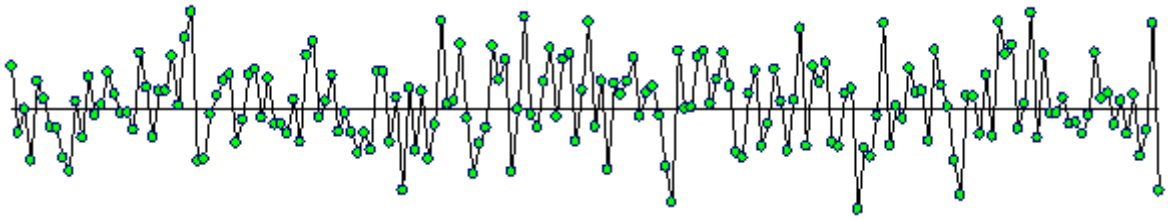


3. Autocorrelation of QC data

Everybody in the field of medical laboratories heard of the Gaussian distribution which is a condition of validity for the quality control theory. Practically this condition is not very demanding because it is generally met, at least approximately with QC data. On the other hand very few technicians are aware of the independence condition which is much more important and very often violated resulting in a heavy deterioration of the performances of QC charts.

Independent time series

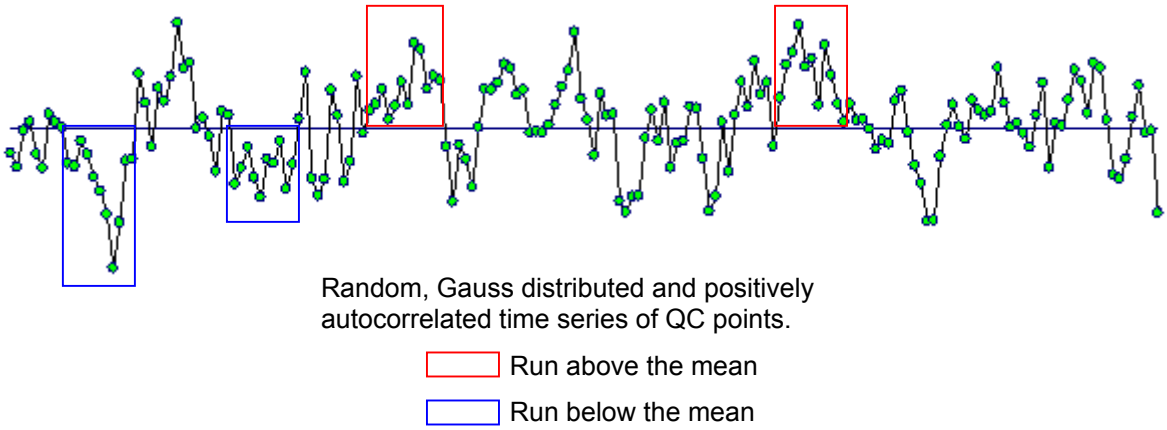
The points in the QC chart below are random, Gauss distributed and mutually independent. The previous observations are not linked to future observations. Successive QC points are uncorrelated. There is no memory in the data. When a point is above the mean, the following one has an equal probability to fall above or below the mean. The process fluctuates randomly around the mean.



Random, Gauss distributed and mutually independent time series of QC points.

Positively auto correlated time series

The points in the chart below are also random and Gauss distributed but positively auto correlated. The most striking feature is the presence of long runs above and below the mean. If the current observation is on one side of the mean, the next observation will be found most likely on the same side of the mean. There is a positive correlation between successive points.



Random, Gauss distributed and positively autocorrelated time series of QC points.

Run above the mean

Run below the mean

Autocorrelation in medical laboratory QC

Skilled clinical chemists have surely recognized the model of many of their QC charts in the positively auto correlated time series. The most important autocorrelating factor of analytical methods is recurrent calibration. Due to the uncertainty of calibration processes (refer to section below), all of the QC points that depends on the same calibration share a common bias which explains their likeness and hence the autocorrelation.

Additionally to recurrent calibration, any external factor of long-term variability is also strongly autocorrelating. For instance aging of reagents, opened on the tray of the analyser, may be a source of autocorrelation because an older reagent may have drifted and thus create a temporary bias which did not exist when the reagent was younger and which will disappear with the next new bottle.

Adverse consequences of autocorrelation on QC performance

The natural occurrence of long runs above and below the mean in autocorrelated time series makes some « control rules » a source of numerous false rejections. This is the case of the well-known

Westgard's rules 2_{2s}, 4_{1s} and 10_x or their association in « multirules ». Two, four or ten QC points plotted at a given distance on the same side of the mean is not necessarily an out-of-control situation when data is auto correlated. So, these Westgard's rules may be as noxious as the 2SD control interval mentioned in a section above. Moreover, the theoretical performance of regular control charts as measured in terms of ARL (average run length) is also damaged by autocorrelation.

From a practical point of view, any theoretically valid QC protocol must not be taken too seriously as long as a strong evidence of efficiency has not been demonstrated at the bench.

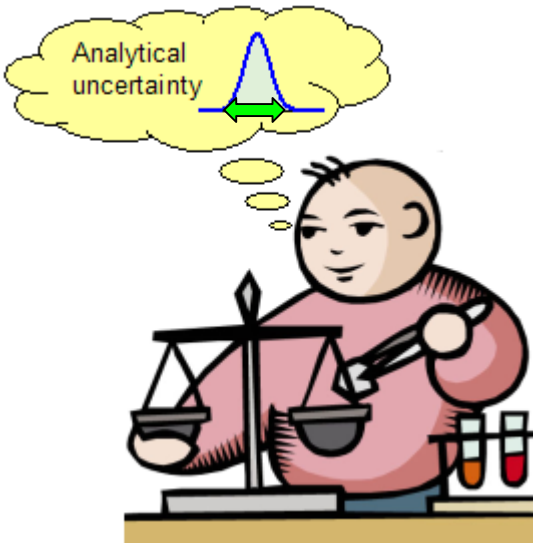
4. QC driven by medical tolerance

Medical tolerance

Making out a table of the acceptable analytical deviations which do not alter diagnosis, follow-up or treatment of patients is the cornerstone on which clinical laboratorty QC must be built. For each assay, the permissible variations of the measured concentrations around the true values define the frontier between conforming and non-conforming.

Some authors speak of « allowed error ». This wording is disputable: An error which is allowed is no longer an error. It is the reason why we borrow from the engineering world the better word « tolerance » which denotes the permissible deviation of an actual property of a product to the par value of this property.

Thus medical tolerance must be explicitly drawn up to provide analysts with a frame to stay in to ensure the clinical usefulness and the marketable value of their assays.



Analytical uncertainty

Any laboratory assay is spoiled by an inherent uncertainty. Random deviations of results are Gauss distributed and thus theoretically boundless. The uncertainty interval is however conventionnally limited to a 6σ spread (m ± 3σ), letting apart extremely rare deviations, the frequency of which is less than 1/370.

Contract between laboratory customers and analysts

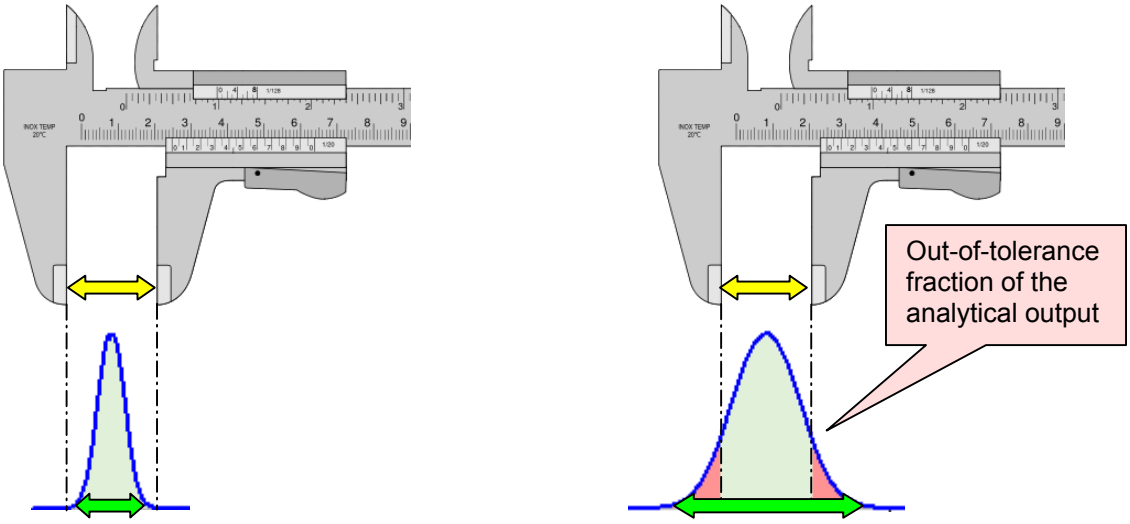
Tolerance specified by medical customers allows analysts a reasonable leeway for inherent variability. It is up to the latters to choose laboratory methods, the uncertainty interval of which does not exceed the tolerance interval. The final aim of quality control is to continuously keep analytical results within-tolerance and so to achieve a zero-defect analytical output (actually less than 1/370 non-confroming assay).

Capability indices

Capability relates tolerance (spread allowed by medical customers) to uncertainty (unavoidable spread of analytical method). The indice C_p is the ratio of the former to the latter:

$$C_p = \frac{Customer's\ tolerance}{Analytical\ \ uncertainty}$$

A necessary (but not sufficient) condition to fulfil the contract between laboratory customers and analysts is thus $C_p > 1$.



Capable process : $C_p = \frac{\leftrightarrows}{\leftrightarrows} > 1$

Uncertainty < Tolerance
Zero-defect analytical output is possible

Incapable process : $C_p = \frac{\leftrightarrows}{\leftrightarrows} < 1$

Uncertainty > Tolerance
A fraction of the analytical output is out-of-tolerance

Exercise 1

Among the two following processes "animal through the hole", which one has a capability C_p greater than 1 ?

Answer : the mouse. Obvious, is not it.

Exercise 2

You purchased a POCT analyser with a sodium direct potentiometric electrode. The long-term CV is, say to be very kind, 1.5%. Is your analyser a cat or a mouse when compared to the hole of medical tolerance for plasma sodium (± 3 mmol/l) ?

Answer : For an average concentration of 140 mmol/l, the 3σ uncertainty half-interval is $3 \times \frac{1.5}{100} \times 140 = 6.3$ mmol/l

The uncertainty interval of your POCT device (± 6.3 mmol/l) is two times larger than the tolerance interval (± 3 mmol/l). You are trying to push a cat through a mouse hole.

5. The uncertainty of calibrations

It is often recommended by manufacturers of clinical chemistry analysers to duplicate calibration assays. Duplication does not kill variation. The variation is only reduced by 30%. Thus every calibration is affected by a great uncertainty.

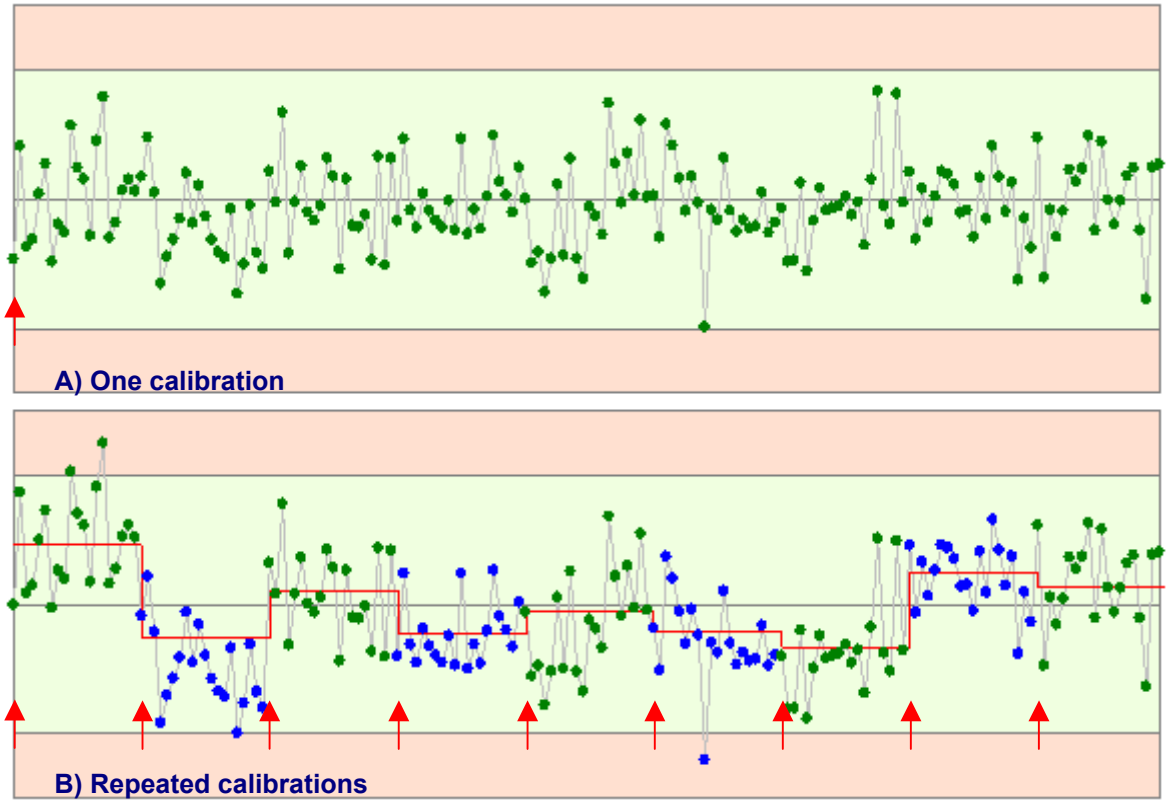
The picture (A) below shows the Shewhart control chart of a stable analytical method which was calibrated only once on the first day.

The picture (B) shows a simulation of what would happen if the method was re-calibrated every twenty days.

The consequences of repeated calibrations are visible on the plot:

- Increased variation
- Strong autocorrelation (refer to section 3)
- Every new calibration creates a different bias which appears as a random shift on the chart
- Higher rate of false rejections

Unnecessary calibrations deteriorate the quality of analytical methods because of the uncertainty of the calibration process. A true metrological calibration would need many more replicates. But this would be economically unacceptable.



A) Shewhart control chart of a stable analytical method calibrated once on the first day.
B) What would happen if the method was re-calibrated every 20 days.

Calibration

Series of shifts brought about by the uncertainty of calibration

Calibrations, as currently performed in clinical chemistry, are lotteries that set the working points of analytical methods within an unforeseeable range of ± 2 CV around the true targets.

Why do most manufacturers of clinical chemistry analysers make it compulsory for their users to perform frequent systematic calibrations ?

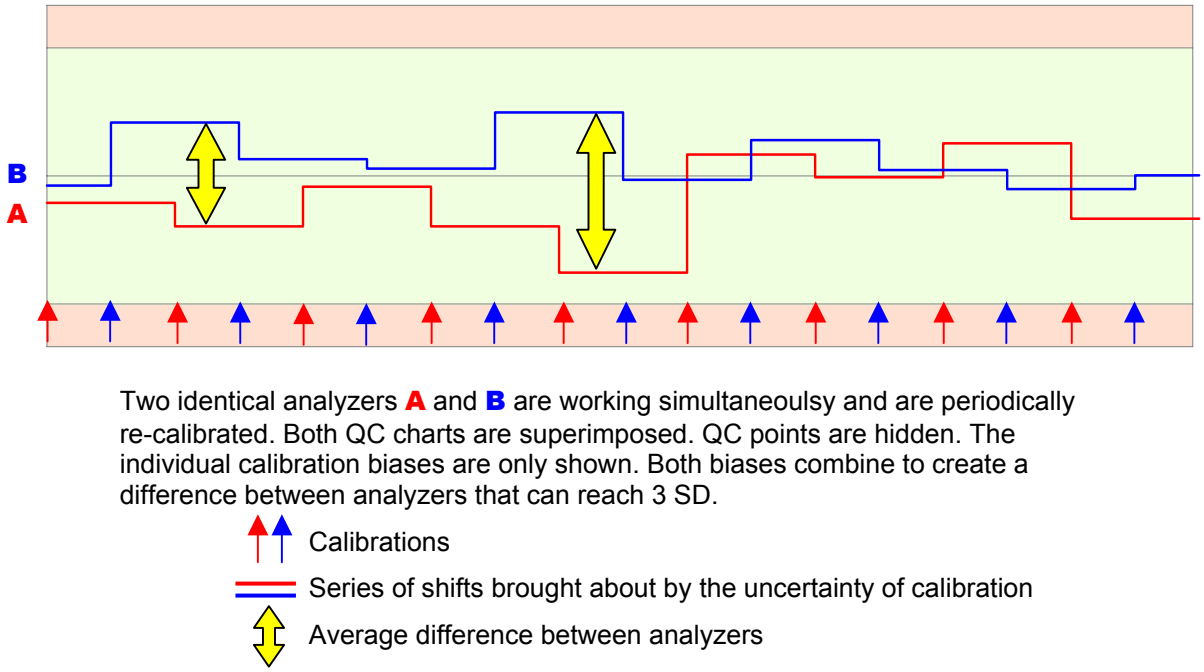
Besides an obvious commercial interest, the alleged claimed reason is "quality". This is perhaps true when users are unskilled operators who might thus be protected against completely wrong analytical results. But this policy for a guaranteed poor quality is unacceptable by skilled operators who are angry and frustrated to see their perfectly stable QC charts ruined by unjustified calibrations that they are compelled to perform.

6. Equalizing two identical analysers

Failure of the metrological approach

Clinical chemistry analysers are often doubled in hospital laboratories to guarantee a 24/24 service in spite of maintenance and breakdowns. Samples are randomly assayed on one of the two instruments, according to the time of the day. So successive samples from an ICU patient can be assayed on any analyser. The agreement must therefore be perfect. This is not automatically ensured even though analysers come from the same manufacturer and are running with the same reagents, calibrators and control materials.

Because of the uncertainty of calibration (refer to section 5), results from each analyzer are independently biased. It is easy to demonstrate that the difference between the two analyzers may range up to 3 SD when calibration assays are duplicated. The picture below illustrates this random discrepancy.



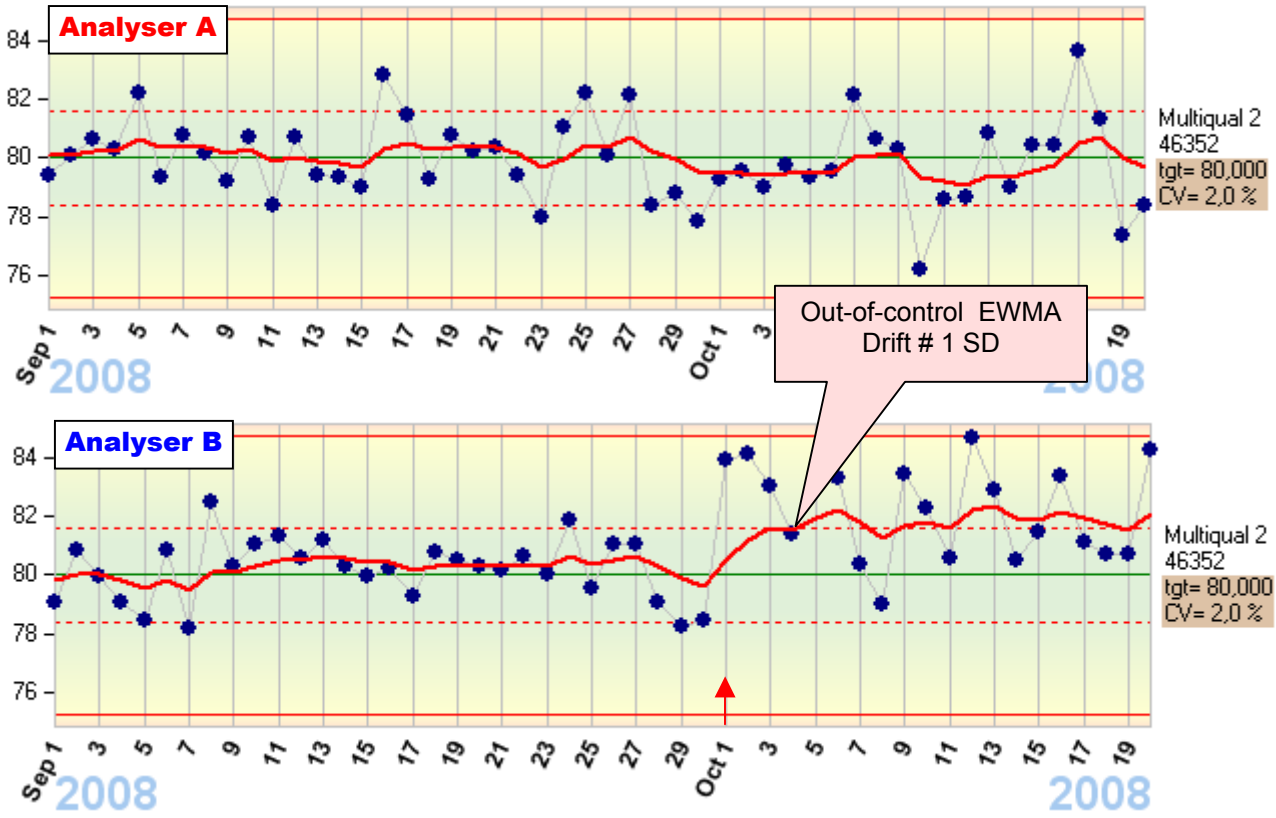
The unavoidable difference between analyzers may be sometimes acceptable if the capability of the analytical method is very high (see section 4). Most often we are compelled to admit that a purely metrological approach cannot meet our needs. The response curves of currently marketted analyzers cannot be *a priori* set with enough accuracy. We must therefore pragmatically turn ourselves to an *a posteriori* regulation which is the only way to master the uncertainty of calibrations.

A touch of engineering process control

Our problem is to equalize, as much as possible, the average output of two analyzers. The solution is easy for anyone who is using a QC software plotting the EWMA (Exponentially Weighted Moving Average) which is a real-time estimate of the average output of analyzers for the different QC levels.

The picture below shows the QC charts of two identical analyzers for the same control material. By chance, the analyzers A and B were perfectly equalized in September. Both EWMA curves were fluctuating around the same mean (80 mg/l).

A re-calibration of analyzer B was performed on Oct 1st. An out-of-control signal occured in analyzer B four days later because of the inertia of the EWMA. The average output of the analyser B could then be estimated as 82 mg/l (about 1 SD too high).



Two analysers from the same manufacturer and running with the same reagents are QCed with the same control material.

Re-calibration

EWMA

EWMA control limits

It would be stupid to consider re-equalizing the analyzers through another calibration. This might equally either improve or worsen the situation. Moreover it would be necessary to wait for additional days to evaluate the new agreement, and to proceed again until we get a satisfying EWMA. Calibration is obviously a too rough procedure to equalize our two analyzers.

What we need is a possibility to manually and finely tune the coefficients of the instrument in an engineering process control fashion. This is particularly desirable since the drift of the EWMA provides us with the precise value of the necessary adjustment.

Practical difficulties

Software of clinical chemistry analysers always have entry fields named *Slope* and *Intercept*. These entries cannot help us because they are intended for a definitive change of analytical methods. They are used for adjustment factors that create permanent shifts. The best example is the compensation of creatinine Jaffe assay to match the ID-MS method.

What we need is a direct access to the calibration factors to be able to slightly adjust them when the EWMA shows that a feedback action is necessary. This adjustement is provisional. It must disappear with the next calibration.

Unfortunately, direct access to calibration factors is rarely available on medical laboratories instruments. As previously mentionned (section 5), this choice is made by manufacturers to guarantee a floor level of (poor) analytical quality with unskilled operators. If you are a skilled and perfectionist clinical chemist you have to purchase another instrument. Choose a new one that allows the feedback adjustement of response curves.

To be followed .....