Internal quality control: planning and implementation strategies

James O Westgard

Address
University of Wisconsin Medical School
1300 University Avenue
Madison, WI 53706, USA

Correspondence
E-mail: jowestgard@med.wisc.edu

This article was prepared at the invitation of the Clinical Sciences Reviews Committee of the Association of Clinical Biochemists.

Abstract

The first essential in setting up internal quality control (IQC) of a test procedure in the clinical laboratory is to select the proper IQC procedure to implement, i.e. choosing the statistical criteria or control rules, and the number of control measurements, according to the quality required for the test and the observed performance of the method. Then the right IQC procedure must be properly implemented. This review focuses on strategies for planning and implementing IQC procedures in order to improve the quality of the IQC. A quantitative planning process is described that can be implemented with graphical tools such as power function or critical-error graphs and charts of operating specifications. Finally, a total QC strategy is formulated to minimize cost and maximize quality.

A general strategy for IQC implementation is recommended that employs a three-stage design in which the first stage provides high error detection, the second stage low false rejection and the third stage prescribes the length of the analytical run, making use of an algorithm involving the average of normal patients' data.

Ann Clin Biochem 2003; **40**: 593–611

Introduction

Healthcare laboratories have a long history of using statistical internal quality control (IQC), beginning with the adaptation of Shewhart's industrial control procedures[1] to the clinical laboratory by Levey and Jennings[2] in 1950. Despite this long period of development, IQC has not matured into a well-developed practice. If anything, IQC has regressed in recent years as the cost of healthcare has taken on a higher priority than quality. QC today often means 'quality compliance', i.e. compliance with regulatory rules, accreditation guidelines and international standards. The weakness of current practice becomes apparent if you ask: What is the quality that is being controlled? How can quality be managed if we don't know the quality that needs to be achieved?

The importance of defining quality requirements becomes clear when analytical quality management is understood in terms of an error budget.[3] To manage any budget, one has to know the amount available to spend. For analytical quality, this means the amount of allowable error. Different sources or components of error can be evaluated and compared with the allowable amount. This can be applied to the initial evaluation of a method[4,5] and extended to ongoing assessment of method performance.[6] If the quality that is required is not known, current laboratory practice is best described not as quality control but as process control, or even 'arbitrary control'. This review describes how arbitrary process control can evolve into true quality control.

Perspective and scope of review

A reference point for this review is Whitehead's 1977 book *Quality Control in Clinical Chemistry*.[7] In that year my own studies[8] on the performance of IQC procedures began at the University of Uppsala. Whitehead outlined the structure needed to control the analytical quality of laboratory tests, which included the following 'QC techniques':

- Stage 1: optimal conditions variance.
- Stage 2: routine conditions variance (known values).
- Stage 3: routine conditions variance (unknown values).
- Stage 4: statistical calculation of patients' results.
- Stage 5: inter-laboratory comparisons.

That structure is still applicable today and will help focus this review.

Stage 1, optimal conditions variance, has to do with assessing the optimal performance of a method and demonstrating that IQC is applicable. This stage has been described by Wernimont[9] as demonstrating that a measurement process is in a state of statistical control. If the method is not able to reproduce results under ideal conditions, its variation is not sufficiently predictable to apply statistical QC. Given the pre-market testing required for the highly automated analytical systems in use today, it is reasonable to expect that most current clinical chemistry methods are capable of being monitored by statistical QC procedures.

Stages 2 and 3 relate to the use of stable control materials, whereas stage 4 makes use of patients' test results. This review will focus on IQC using stable control materials, with occasional extensions to QC procedures on patients' data.

Finally, stage 5 deals with external quality assessment ('peer comparison programmes', 'proficiency testing programmes'). The main value of these programmes is to provide an estimate of the bias observed between an individual laboratory's result and that of its reference or peer group.

I should emphasize that my views reflect my experience as a clinical chemist working in a US hospital laboratory, where QC practices are strongly influenced by government regulations. Today is not the golden age of quality in healthcare or in healthcare laboratories. We can and should be doing better.

Basic principles and terminology

The fundamental principle of IQC is to compare process performance with what is expected under stable operation, as illustrated in Fig. 1. Stable performance is first established by assaying control materials over a period of time, then calculating the mean and standard deviation (SD or *s*). Measurements are then continually made on those same control materials and compared with the original distribution, usually by plotting them on control charts that have control limits set as the mean, plus and minus certain multiples of the SD (often 2*s* and/or 3*s*). Unexpected values

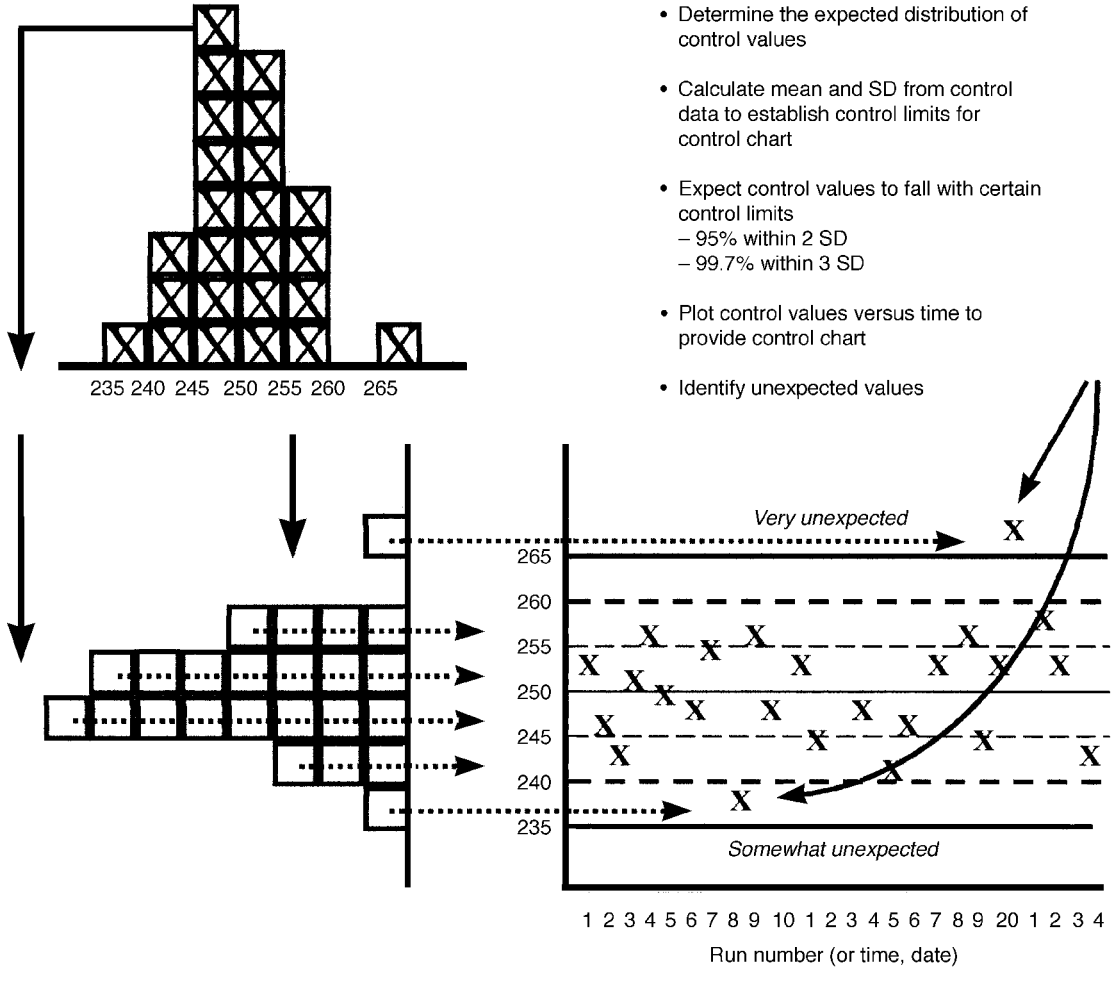

Figure 1. *Conceptual basis of the control chart. SD, standard deviation.*

are identified to alert the analyst to possible changes in process performance.

A **control chart** is a graphical method for displaying control results. Control results are usually plotted against time or sequential run number. Lines are often drawn from point to point to accent any trends, systematic shifts, or random excursions. In healthcare laboratory applications, where the practice is to plot individual control rules, the control chart is commonly known as a **Levey–Jennings chart**, even though the use of individual control values was actually introduced by Henry and Segalove.[10]

The **analytical run** refers to the interval – period of time or group of patients' samples – for which a decision on control status is to be made. The length of an analytical run varies from system to system and laboratory to laboratory, depending on the stability of the analytical system and its susceptibility to changes, such as operators, reagents, recalibration, or other factors that may introduce problems.

Control limits are lines drawn on a control chart to provide graphical criteria for assessing whether a measurement procedure is in or out of control. These limits are usually calculated from the mean and standard deviation which were determined under stable operation.

Control rule means a decision criterion for judging whether an analytical run is in or out of control. It is often represented[8] by a symbol of the form A_L, where A is an abbreviation for a statistic or represents a number of control measurements, and L identifies the control limits. Thus:

- 1_{3s} refers to a control rule that is used with a Levey–Jennings chart when the control limits are set as the mean $\pm 3s$. A run is rejected when a single control measurement exceeds either control limit.
- 1_{2s} refers to the control rule when the control limits are set as the mean $\pm 2s$. Shewhart[1] always considered $2s$ limits to be 'warning' limits, but healthcare laboratories often use $2s$ limits for run rejection.
- 2_{2s} refers to the control rule where a run is rejected when two consecutive control measurements exceed the same mean+$2s$ limit or the same mean $-2s$ limit.
- R_{4s} refers to a control rule where a run is rejected when one control measurement in a group exceeds the mean+$2s$ limit and another exceeds the mean $-2s$ limit.
- 4_{1s} is a rule where a run is rejected when four consecutive measurements exceed the same mean+$1s$ or the same mean $-1s$.
- $10x$ is a rule where a run is rejected when 10 consecutive measurements fall on one side of the mean. There are other rules that vary the number of measurements from 7 to 12, e.g. $7x$, $8x$, $9x$ and $12x$.

Multi-rule QC refers to the use of a combination of control rules for interpreting control data. For example, the combination $1_{3s}/2_{2s}/R_{4s}/4_{1s}/10x$ was recommended[11] for clinical chemistry applications in the early 1980s when IQC mainly used either the 1_{2s} or 1_{3s} rule. This multi-rule application, shown in Fig. 2 and commonly known as '**Westgard rules**', made use of a 1_{2s} rule as a warning to trigger inspection by two rules that are sensitive to random error (1_{3s} and R_{4s}) and three rules that are sensitive to systematic error (2_{2s}, 4_{1s} and $10x$). The objectives were to reduce the false rejections that occurred when $2s$ limits were used for rejection, to improve error detection over that achieved when $3s$ limits were used, and provide guidance for problem-solving by relating the control rule violated to the type of error that would be expected.

Cumulative sum QC, or **cusum**, refers to a control technique in which the difference between the observed control values and the expected mean is calculated, then summed to provide a more sensitive

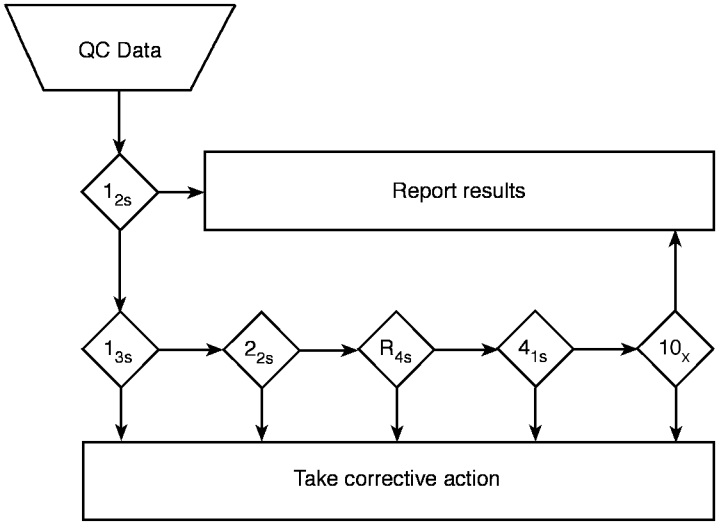

Figure 2. *Flowchart and logic for the multi-rule internal quality control (IQC) procedure commonly known as 'Westgard rules'. See text for definition of symbols.*

indicator of systematic changes.[12] For example, if several control results fall above the mean, all the differences are positive and the cumulative sum of those differences becomes larger and larger with each additional point that is above the mean. Cusum has been widely used in industrial applications[13] and has been strongly recommended for healthcare laboratory applications.[14]

Mean and range QC refers to the classical Shewhart procedures[1] in which the mean and range of a group are plotted on two different control charts. Mean and range QC is widely used in industry, but has not been widely used in healthcare laboratories, which have instead used the mean of patients' data for QC purposes. This application is often known[15] as '**average of normals**' and usually involves calculation of the mean of a truncated patient distribution that requires tens to hundreds of patients' results to monitor analytical performance, depending on the test in question and the ratio of the population variation to the analytical variation.[16]

Planning strategies (doing the right IQC)

Quality is often described as 'doing the right thing right'. The quality of IQC depends on 'doing the right IQC right'. The first right implies performing IQC with the correct number of control measurements and applying the correct statistical control rules. The second right refers to implementing IQC correctly by selecting appropriate control materials, calculating the control data properly, setting control limits correctly, interpreting the control data correctly and responding to control signals properly. Doing the right IQC has to do with planning and design. Doing QC right has to do with implementing the QC design properly.

General guidelines for planning and design of IQC procedures have been provided by NCCLS (National Committee for Clinical Laboratory Standards).[17] The steps for planning a statistical QC procedure are presented as follows:

1 Define the quality requirement for the test.
2 Determine method precision and bias.
3 Identify candidate IQC procedures.
4 Predict IQC performance.
5 Set goals for IQC performance.
6 Select an appropriate IQC procedure.

While the NCCLS document provides this general framework, it does not specify details on how to perform this planning process. Here's a more detailed discussion of the steps, including a simple example to illustrate the process.

1. Define the quality requirement for the test

Many different ways of defining quality requirements for laboratory tests have been recommended, including the allowable total error (TE_a) from proficiency testing (PT) and external quality assessment (EQA) programmes, maximum allowable SD (SD_{max}) and maximum allowable bias ($bias_{max}$) based on biological variability, and medically or clinically important changes (clinical decision interval, D_{int}) based on test interpretation guidelines. In 1999, a consensus conference recommended[18] a hierarchy of quality requirements that includes all these different formats, giving preference to clinical over biological over PT/EQA over expert group recommendations over state-of-the-art performance figures.

Figure 3 shows my perspective on these different types of quality criteria and their relationship to the operating specifications that describe the imprecision (s_{meas}) and inaccuracy ($bias_{meas}$) allowable for a method and the IQC procedure needed (control rules and number of control measurements, N). Starting at the top of the figure, medically important changes in test results can be defined by standard treatment

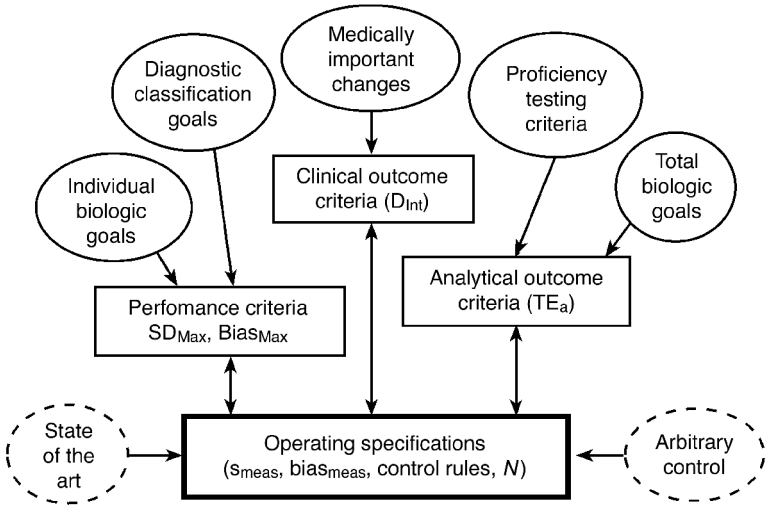

Figure 3. *A system of quality requirements and operating specifications. See text for definition of symbols.*

guidelines (clinical pathways, clinical practice guidelines, etc.) to establish clinical outcome criteria (or decision intervals, D_{int}). Such clinical criteria can be converted to laboratory operating specifications for imprecision (s_{meas}), inaccuracy ($bias_{meas}$) and QC (control rules, N) by a clinical quality planning model[19] that takes into account pre-analytical factors, such as individual or within-subject biological variation. Note that some earlier models did not account for within-subject biological variability, and the recommendations for medically allowable standard deviations were therefore erroneously large.

The left side of the figure shows how performance criteria for imprecision and inaccuracy can be defined as separate analytical goals for the maximum imprecision and bias that would be allowable for the stable performance of the method. Specifications for maximum imprecision can be formulated on the basis of within-subject biological variation.[20] The maximum allowable bias can be derived from diagnostic classification models, as described by Klee.[21]

The right side of the figure shows how proficiency testing criteria define analytical outcome criteria in the form of TE_a, which can likewise be translated into operating specifications (s_{meas}, $bias_{meas}$, control rules, N) via an analytical quality planning model.[22] Note that the TE_a can also be set on the basis of total biological goals that are population-based or individual-based,[20] so that the extensive data bank of individual biological variation can also be utilized in this situation.[23]

2. Determine method precision and bias

Estimates may be obtained from the initial validation studies performed on new analytical methods, e.g. precision from a replication experiment and bias from a comparison of methods. With established methods, the estimate of precision can be obtained from routine IQC data, and the estimate of bias from periodic EQA or PT results.

3. Identify candidate IQC procedures

This step is concerned with the types of control rules that can be implemented in the laboratory. These candidate procedures will depend on preferences in the local or national marketplace and the IQC programmes and software that are available. Laboratories typically prefer single-value or individual-value control procedures (Levey–Jennings type of IQC) to mean/range and cusum procedures. However, there are situations where cusum, mean/range and average of normal patients' data procedures are preferred.

4. Predict IQC performance

Closely related to the selection of candidate IQC procedures is the availability of information about the performance of those procedures. Performance figures for IQC procedures are generally provided in terms of the probability of rejection or the average run length. Figure 4 shows the 'power curves' for common IQC procedures having two control measurements per run. Note the two situations of interest:

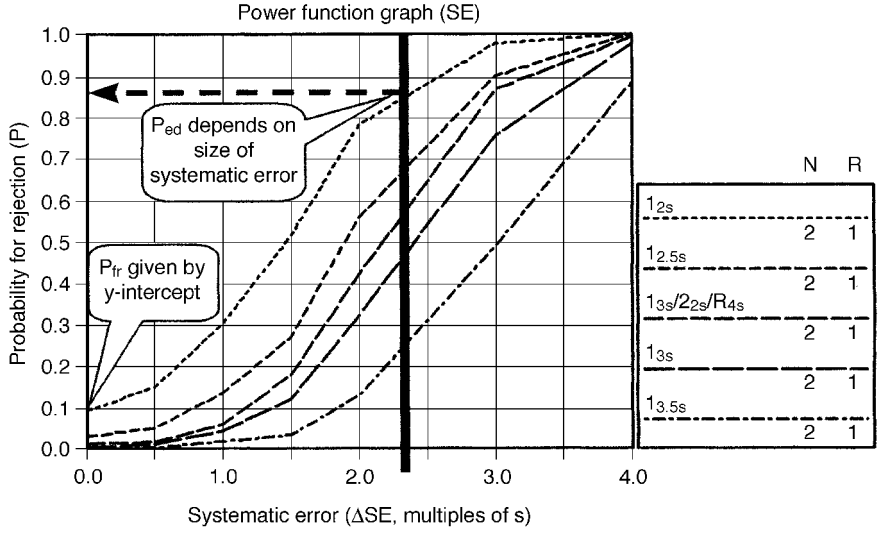

Figure 4. *Power curves for commonly used internal quality control (IQC) procedures having two control measurements per run. The probability of rejecting an analytical run is shown on the* y*-axis versus the size of systematic error on the* x*-axis. The individual curves correspond to the control rules and numbers of control measurements shown in the key at the right.* P_{fr}*, the probability for false rejection, is given by the* y*-intercept.* P_{ed}*, the probability for error detection, depends on the size error occurring, such as illustrated by the vertical line, and estimated by reading the* y*-value at the point of intersection with the power curve. See text for definition of symbols.*

- Performance when there are no errors occurring except the stable random analytical variability of the method or process, as described by the probability for false rejection (P_{fr}) or the average run length for acceptable quality (ARL_{accept}). Ideally, P_{fr} is close to zero and the ARL_{accept} is very long, meaning that there will be few false rejections and there will be a long time between them.
- Performance when an additional error is occurring, such as a systematic shift or an increase in the method's SD, as described by the probability for error detection (P_{ed}) or the average run length for rejectable quality (ARL_{reject}). Ideally, both P_{ed} and ARL_{reject} should be close to unity, meaning there will be a high chance of detecting errors, preferably within the first run in which they appear.

These probability and ARL terms are related, and ARL can be calculated as the reciprocal of the corresponding probability term (i.e. $ARL_{accept} = 1/P_{fr}$ and $ARL_{reject} = 1/P_{ed}$) when the probability characteristic is constant from run to run.[24] The calculation becomes more complicated when the probability changes from run to run, as happens with cusum and multi-rule procedures. The earliest and most widespread performance data in the clinical chemistry literature make use of probabilities for rejection,[11,25] but some investigators[26,27] strongly recommend the use of ARLs. Furthermore, ARLs can be converted to units of time to characterize the average time until a rejection occurs, as illustrated in a published validation study of a new IQC technology.[28]

5. Set goals for IQC performance

The ideal probability for false rejection would be 0.000, which would mean that a false rejection would never occur (ARL_{accept} = infinity). In practice, I would recommend that the maximum P_{fr} be 5%, that 1% or less is a more desirable goal and that 0.1% is even better (ARLs of 20, 100 and 1000, respectively). The immediate implication is that the use of 2*s* control limits should be avoided since the false rejection probability is about 4.5% when there is one control measurement per run, 9% for two per run, 14% for three per run and 18% for four per run. The use of 3*s* control limits with *Ns* from 1 to 6 will keep false rejections at 1% or less. The use of 3.5*s* control limits with *Ns* from 1 to 4 will keep false rejection at 0.1% or less.

The ideal probability for error detection would be 1.000, which would mean that the IQC procedure would always detect the error. The analysis of this situation is more complicated because P_{ed} depends on the size of the error occurring, as shown by the power function graph[27] in Fig. 5. The real objective here should be to achieve a high P_{ed} for medically important errors, e.g. a probability of 0.90 or a 90% chance of detecting a critical systematic shift that would cause a 5% risk of incorrect test results. This critical systematic error (ΔSE_{crit}) can be calculated once the quality requirement is defined and the method precision and bias are known, using the following equation:

$$\Delta SE_{crit} = [(TE_a - bias_{meas})/s_{meas}] - 1.65$$

where TE_a is the quality requirement in terms of a TE_a, $bias_{meas}$ is the method bias, s_{meas} is the method precision and 1.65 is the *z*-value that corresponds to 5% of the area in one tail of a Gaussian distribution.

Note that Linnet adopted a similar definition for the 'maximum clinically allowable analytical error' in his methodology for choosing IQC procedures.[29] Six sigma quality management also includes a capability index that can be directly related[30] to ΔSE_{crit}, i.e. sigma capability metric $= (TE_a - bias_{meas})/s_{meas} = \Delta SE_{crit} + 1.65s$, thus the IQC procedure depends directly on the capability of the measurement procedure.

As a sample calculation, consider that the TE_a for a cholesterol method is 10% [given by the US Clinical Laboratory Improvement Amendments (US CLIA) regulations as the criterion for acceptability of proficiency testing results]. For a method that has a coefficient of variation (CV) of 2.0% and a bias of 1.0%, ΔSE_{crit} would be equivalent to $2.85s\{[(10-1)/2]-1.65\}$. The vertical line in Fig. 5 shows this size of systematic error imposed on the power curves of several IQC procedures with *Ns* from 1 to 6. The points of intersection with the power curves show that the probabilities of error detection will vary from 0.40 to 1.00, depending on the choice of control rules and the number of control measurements. As would be expected, the more control measurements, the better the error detection; the use of multi-rule procedures also gives better error detection than single-rule procedures having the same number of control measurements.

6. Select an appropriate IQC procedure

The right IQC procedure is one that has at least a 0.90 probability or 90% chance of detecting medically important errors ($ARL_{reject} = 1.1$) and a maximum 0.05 probability or 5% chance of false rejections ($ARL_{accept} = 20$), preferably 1% or less ($ARL_{accept} = 100$ or more). Simple control rules are preferred over multi-rule combinations or more complicated cusum calculations, along with the lowest number of control measurements that will provide the desired error detection.

For the cholesterol example cited above, the ΔSE_{crit} should be detected over 90% of the time using four control measurements with either a $1_{3s}/2_{2s}/R_{4s}/4_{1s}$ multi-rule procedure or a $1_{2.5s}$ single-rule procedure.

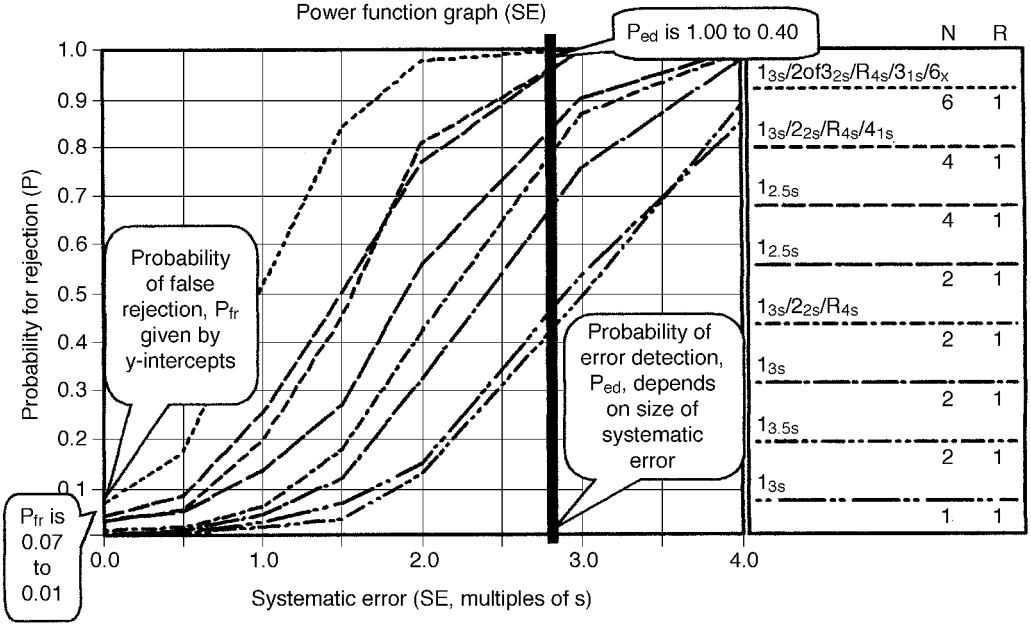

Figure 5. *Example of the use of a power function graph to assess internal quality control (IQC) performance for a cholesterol test where the medically important systematic error is equivalent to 2.85 times the standard deviation of the method. See text for definition of symbols.*

Both have the desired 90% error detection with false rejection less than 5%. Both are appropriate IQC designs for this application. IQC procedures with $N = 2$ will not provide the necessary error detection. IQC procedures with $N = 6$ are wasteful of laboratory resources.

Recommendation for a practical planning tool

While the principles of IQC planning have been discussed in the clinical chemistry literature for almost 25 years, laboratories have been slow to implement an IQC planning process. Many laboratory analysts are put off by statistics and are slow to make changes in IQC. That is due in part to the lack of an IQC planning or selection requirement in existing regulations, accreditation guidelines and international standards, but also in part to the lack of practical tools that make the selection of IQC procedures quick and easy. Early planning processes depended on QC simulation programmes,[31] which have never been widely available. While simulation and modelling are crucial for determining the performance characteristics of IQC procedures and understanding the IQC process,[32] these tools have never become practical in service laboratories. A more practical tool is now available: the chart of operating specifications (OPSpecs chart).[33,34] The advantage of this tool is that it provides a graphical display of the relationship between the quality requirement of the test, the precision and accuracy of the method and the error detection of the IQC procedure.

The OPSpecs chart is based on an equation that is a rearrangement of the critical-error calculation, as follows:

$$\text{bias}_{\text{meas}} = \text{TE}_{\text{a}} - (\Delta\text{SE}_{\text{crit}} + 1.65)s_{\text{meas}}$$

Note that this is an equation for a straight line ($y = a + bx$). When $\text{bias}_{\text{meas}}$ is plotted on the y-axis versus s_{meas} on the x-axis, the y-intercept is TE_{a} and the slope is $\Delta\text{SE}_{\text{crit}} + 1.65$. By specifying the value for P_{ed} as 0.90 or 90%, the size of the $\Delta\text{SE}_{\text{crit}}$ that can be detected by various IQC procedures can be read from the power curves, as shown in Fig. 6. For example, the multi-rule procedure with $N = 6$ can detect a systematic error equivalent to $1.7s$ with 90% error detection (left-most vertical arrow). Compare that to the 1_{3s} rule with $N = 2$, which can only detect an error of $3.6s$ with 90% detection (sixth arrow from the left in Fig. 6). By substituting the size error that can be detected and defining the quality requirement for the test, it is possible to display graphically the bias and imprecision that are allowable for a method and the QC that is necessary to guarantee or assure the quality required for the test.

Figure 7 shows the OPSpecs chart for these same IQC procedures. $\text{Bias}_{\text{meas}}$ is plotted on the y-axis versus s_{meas} on the x-axis. The y-intercept corresponds to the defined quality requirement, in this case a TE_{a} of 10%. The various lines correspond to the different IQC procedures whose rules and Ns

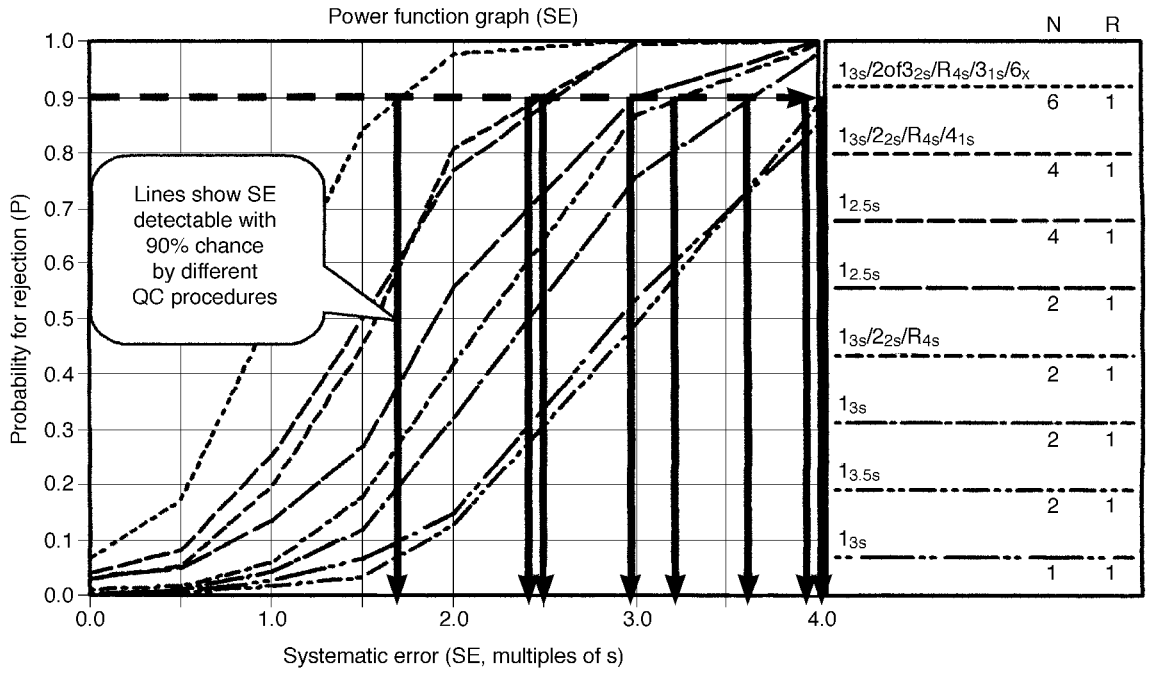

Figure 6. *Sizes of systematic errors that can be detected with 90% confidence by different internal quality control (IQC) procedures. See text for definition of symbols.*

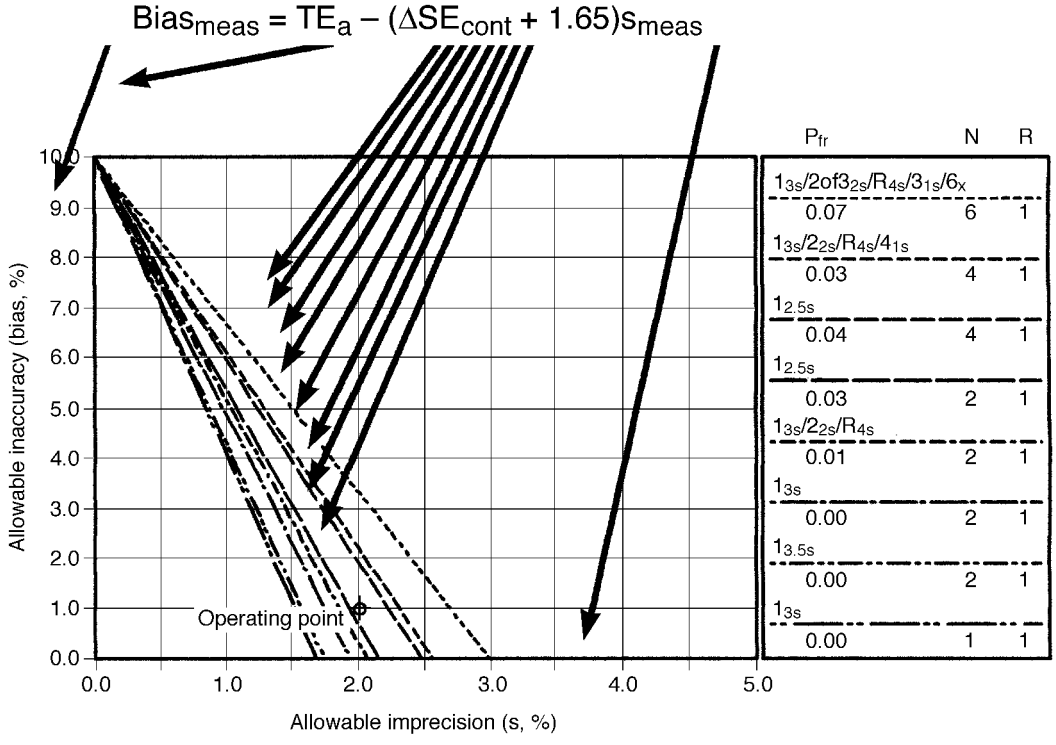

Figure 7. *Mathematical basis for a chart of operating specifications that shows the bias and imprecision that are allowable for different IQC procedures, whose control rules and* Ns *are given in the key at the right. See text for definition of symbols.*

are shown in the key at the right side. The order of the lines – top to bottom – is the same as the order in the key, i.e. the highest line corresponds to a multi-rule procedure with $N=6$ and the bottom line corresponds to a 1_{3s} rule with $N=1$. Each line defines the region of allowable bias and allowable imprecision for a method if the IQC procedure is to provide 90% analytical quality assurance

(AQA) for detection of systematic errors, which is specified on an OPSpecs chart by the notation 90% AQA(SE). Note that OPSpecs charts could be prepared for other levels of AQA, such as 50% AQA(SE) or 25% AQA(SE), as well as for random analytical error.

The OPSpecs chart can be prepared by an electronic spreadsheet[34] or by specific computer programs created for this task (QC Validator and EZ Rules programs from Westgard QC, www.westgard.com). A manual of OPSpecs charts has also been prepared,[35] and 'normalized' OPSpecs charts are available in a book on IQC planning[36] and on the Internet.[37] An advantage of the computer programs is the availability of an automatic selection process that is initiated by input of the number of control materials to be analysed.[38,39]

In using an OPSpecs chart to select an IQC procedure, an 'operating point' is plotted to represent the method's bias (y-coordinate) and precision (x-coordinate). An acceptable IQC procedure is one whose line for allowable bias and allowable precision is above the operating point. For example, Fig. 8 shows an OPSpecs chart for an allowable TE of 10% and IQC procedures that will provide 90% detection of systematic errors. The operating point represents our earlier cholesterol example. The three lines above the operating point represent the acceptable IQC procedures.

Recommendation for a practical planning process

The right IQC procedure can be selected using the process outlined in Fig. 9. This selection or planning process makes use of the normalized OPSpecs charts that are presented in the *Basic Planning for Quality* manual.[36,37]

- The process begins by defining the quality requirement as a TE_a in percentage units. It is usually helpful to select a medically important concentration where test interpretation is critical, define the allowable error, then calculate the percentage at that medical decision concentration.
- Estimates of method bias and precision should also be calculated in percentages.
- The normalized operating point has a y-coordinate given by $\%bias/\%TE_a$ and an x-coordinate given by $\%CV/\%TE_a$.
- The normalized operating point is plotted first on a 90% AQA chart with $N = 2$ or 3, depending on the number of control materials to be analysed.
- The OPSpecs chart is inspected to determine whether any of the candidate IQC procedures are adequate. If no IQC procedure can be selected, the operating point is plotted on a 90% AQA chart having $N = 4$ or 6 (twice the number of control measurements). If no IQC procedure can be

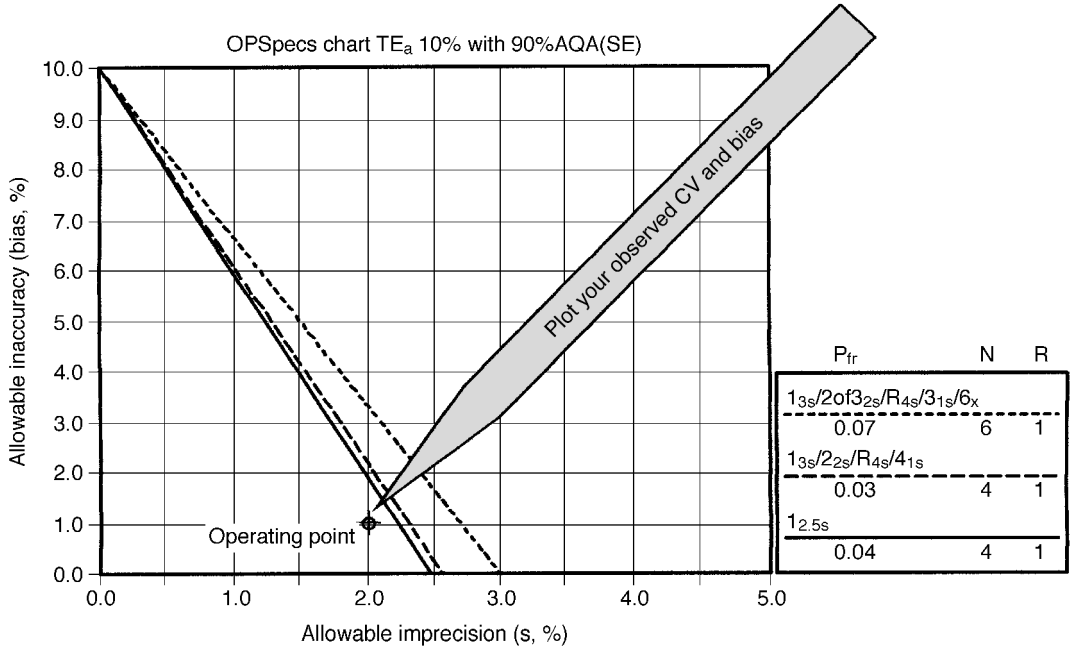

Figure 8. *Application of the OPSpecs tool by plotting the observed method bias as the* y*-coordinate and the observed method precision as the* x*-coordinate to describe the 'operating point' of an analytical method. See text for definition of symbols.*

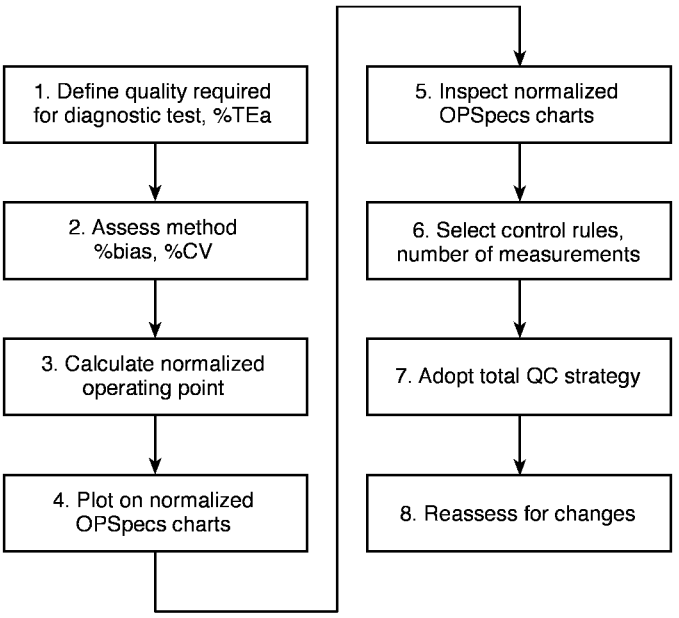

Figure 9. *A practical procedure for planning and selecting internal quality control (IQC) procedures. See text for definition of symbols.*

selected, then the operating point is plotted on a 50% AQA chart having $N = 4$ or 6.

- The control rules and number of control measurements are selected to achieve 90% error detection with the minimum false rejections. Single-rule procedures can be given preference when a solution is found on a 90% AQA chart. Multi-rule procedures should be given preference when the solution is found on a 50% AQA chart.
- A total QC (TQC) strategy is identified on the basis of the error detection that can be expected for a selected IQC procedure, i.e. high probability for error detection (HI-Ped) when the solution is found on a 90% AQA chart, moderate error detection (MOD-Ped) when the solution is found on a 50% AQA chart and low error detection (LO-Ped) when there is no solution on a 50% AQA chart.
- Whenever there is a change in method performance or quality requirement, the IQC design should be reassessed by repeating this process.

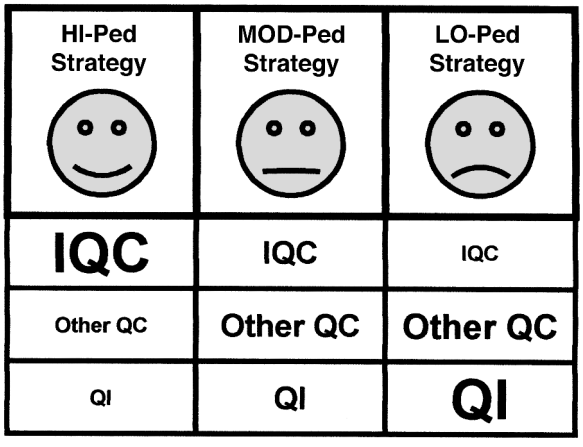

Figure 10. *Total quality control (TQC) strategies that illustrate the proper combination of IQC with instrument and function checks (Other QC) and need for quality improvement (QI). HI-Ped, high probability for error detection; MOD-Ped, moderate probability for error detection; LO-Ped, low probablility for error detection.*

The outcome of this selection process is both an IQC procedure and a TQC strategy that complements the expected performance of the IQC procedure. The idea behind the TQC strategy is illustrated in Fig. 10, where the relative size of the abbreviations represents the relative effort that should be made in IQC, other QC and quality improvement (QI). 'Other QC' refers to other performance tests, instrument checks, etc. QI here emphasizes the reduction of bias and CV to improve method performance. Note that the OPSpecs chart can provide guidance as to the improvements that are needed: simply move the operating point to see what precision and bias would lead to better QC and easier management of the testing process.

Figure 11 provides a detailed flowchart to guide the development of a TQC strategy. The steps on the left apply when 90% error detection can be achieved, the steps in the middle apply for 50% error detection, and the steps at the right apply when less than 50% error detection is available. In short, when 90% detection is achievable, the TQC effort is aimed at doing the minimal IQC necessary using the simplest rules and lowest number of control measurements. When only 50% detection is achievable, the TQC effort maximizes

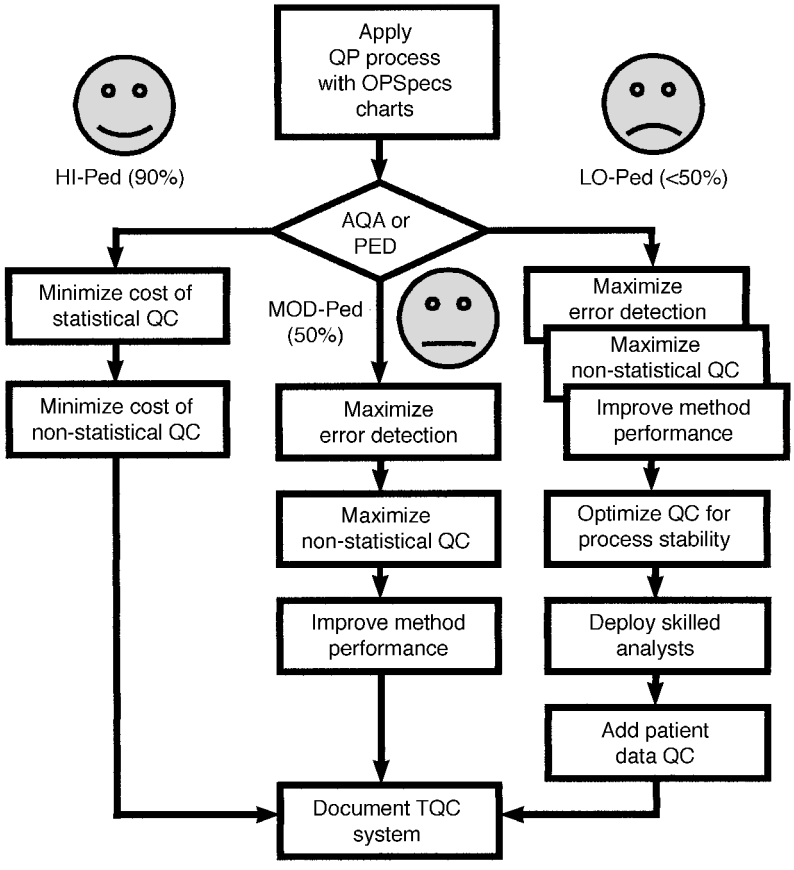

Figure 11. *A detailed flowchart to guide the development of TQC strategies. See text and legend to Fig. 10 for explanation of abbreviations.*

the IQC, often by selecting multiple control rules and higher numbers of control measurements. When detection is less than 50%, the maximum IQC is performed and, in addition, highly trained analysts should be assigned and patients' data should also be inspected. Much more time and effort is required to monitor a process when IQC yields low error detection, and there will therefore be a big advantage if the analytical performance can be improved and the method moved from a low to moderate to high TQC strategy.

Although this quality-planning process may seem complicated, it is quite easy to perform. Once analysts have been trained or have some practice, it takes only a minute or two to identify an appropriate IQC procedure and TQC strategy. The real limitation in practice is the time it takes to define the quality requirement, collect the information about method performance, and document the results. The QC Validator and EZ Rules computer programs take care of the documentation automatically and can also provide a clinical quality planning model that is appropriate for translating medically important changes into operating specifications for precision, accuracy and IQC.[19] In addition, the EZ Rules program includes the performance characteristics for QC procedures based on average of normals (AON) patients' data and can be used to select appropriate numbers of patient samples in AON algorithms.[40]

Implementation strategies (doing IQC right)

There are many factors that are important in the real-world operation of IQC of healthcare laboratories. The operation begins by selection of the right control rules and numbers of control measurements as described above. Some 'good practice' guidelines are immediately apparent:

- DON'T use 2 SD control limits.
- DON'T use the same control rules for all tests.
- DO select IQC procedures for individual tests on the basis of the quality required for the test and the precision and accuracy observed for the method.
- DO minimize false rejections in order to maximize response to real problems when they occur.
- DO build in the error detection necessary to detect medically important errors by selection of appropriate control rules and numbers of control measurements.

- DO complement the IQC procedure with an appropriate TQC strategy.

Some other factors that must be considered are discussed in this section.

Stability of control materials

The principle of IQC is that the variation observed represents primarily the variation of the measurement process. Additional sources of variation, such as bottle-to-bottle variation of control materials and the long-term stability of a batch of control materials, can reduce the error detection capabilities of the IQC procedures. The simple IQC procedures that utilize individual control values assume a single component of variation. Multiple components of variation, e.g. even within-run and between-run components of measurement variation, generally degrade the performance of the IQC procedure.[41] It is therefore critical to obtain stable materials with minimum bottle-to-bottle variation. Liquid control materials may be preferred to reduce this variation, but lyophilized materials may offer better long-term stability. Commercial control materials are expected to have better stability than locally prepared patients' pools, but there are always issues of whether the sources are human or animal and their possible effects on the behaviour of the measurements of interest.

Concentrations of analytes

Materials should also be selected to have concentrations close to those critical for the clinical interpretation of a test. Generally, two or three concentrations levels are recommended in order to monitor performance throughout the analytical range. The desire to utilize only two or three control materials for many different analytes will limit the ability to have all analytes at their desired concentrations.

Calculation of control data and control limits

Given that the fundamental principle of IQC is to compare the variation observed today with that under stable operating conditions, the means and SDs for control materials should be calculated from data obtained in the laboratory. Control limits should not be calculated from values assigned by the manufacturer or peer group values obtained from an EQA programme. Those sources undoubtedly include a between-laboratory component of variation that could be many times larger than the within-laboratory variation. The NCCLS C24-A2 document does recognize that assigned values may be needed when beginning with a new control material, but recommends that the laboratory switch over to its own calculated values as soon as possible (after collection of 20 data points).[17] NCCLS further recommends the use of cumulative means, SDs and control limits that represent 100 or more measurements (up to 6 months of data) in order to minimize the variability in the estimate of the method SD.

Response to out-of-control signals

If the IQC procedure has been properly designed to minimize false rejections, if control materials are stable and primarily reflect the variation due to the method and if control limits are correctly calculated to describe the variation observed within a laboratory, the proper response to an out-of-control signal is to stop the process and determine what is wrong. That recommendation may be the most important part of the NCCLS C24-A2 document.[17] The common practice of simply repeating the control until the measurement finally is 'in control' is wasteful and wrong. However, laboratories are conditioned to this practice because of their historical use of 2 SD control limits and their experience with the false rejections inherent with 2 SD limits. The only way to break out of this bad practice is to design the IQC procedure to minimize false rejections and to implement it correctly to be sure the control limits truly reflect the performance of the method in the laboratory.

Run length, frequency and location of control samples

These are difficult and inter-related issues that are not yet resolved. NCCLS defines run length in terms of the period of time or number of specimens for which the precision and accuracy are expected to be stable.[17] There are two dimensions to this idea of stability: (1) the stable performance documented by the manufacturer during studies under close to ideal conditions; and (2) the susceptibility of the measurement process to changes that might occur. The manufacturer may be able to define the maximum run length, but the laboratory user will almost always need to define a shorter period to account for factors that vary in the laboratory, e.g. changes in staffing and shifts.

The NCCLS definition is flexible and adaptable to different types of analytical systems and different modes of operation, e.g. random-access continuous reporting versus batch-process reporting of results. In random-access operation, it may be desirable to place control materials up front in order to validate the measurement performance prior to the release of any test results, then periodically revalidate performance. In batch processing, control materials may be spread throughout the analytical run to validate performance prior to reporting the whole batch. The key is to recognize when the QC data need to be interpreted and to be sure there are sufficient measurements available at that time to guarantee the quality of the test results.

Parvin[26,27] has recommended the regular placement of controls throughout a run, regardless of the mode of operation, in order to detect large errors as early as possible, not just to detect the minimal size error that would invalidate test quality. The minimal size error should be monitored by 'event-driven' IQC, e.g. at the beginning or start-up of an analytical system, or after change of a component in that system, whereas the regular placement of control samples is considered non-event-driven IQC. Others have applied a survey methodology to define run length better on the basis of experiences with a given analytical system.[42] My own preference would be to utilize patients' data to determine analytical run length as part of a multiple QC design strategy.[40]

Recommendation for general structure of an IQC procedure

The basic structure of an IQC procedure should consider three different designs or stages of operation, as follows:

1 A 'start-up' or event-driven IQC procedure.
2 A 'monitor' or non-event-driven IQC procedure.
3 An average of normals (AON) patients' data QC procedure.

The start-up procedure should provide high error detection to ensure that everything is working correctly at the beginning of an analytical run and before patients' test results are reported. The monitoring procedure should yield low false rejection rates so that repeat analyses and waste of time and effort are avoided. The AON procedure should monitor the stability of the analytical run to determine when the process needs to be revalidated.

In this generic model, the start-up and monitor designs could be the same if the goals for high error detection and low false rejection can be achieved with a single design. This is possible for highly automated analysers, such as the fourth- and fifth-generation chemistry and haematology systems. For earlier generations of automated systems, such as those for immunoassays, it is more likely that the start-up design will need a multi-rule procedure with a high number of control measurements in order to achieve the desired error detection, but the monitor design can utilize a single-rule procedure with lower N.

The concept of multi-stage IQC is not new, having first been recommended for clinical chemistry applications almost 20 years ago.[43] However, it has not been widely implemented because of the need for development and improvements in the IQC software available to laboratories. One sample application was described in 1987 by Eggert *et al.*,[44] who developed a multi-rule multi-stage IQC programme for a laboratory information system. Today, multi-stage designs can be implemented by IQC programmes on personal computers.[45]

A sample implementation of multi-stage IQC is shown in Fig. 12. Note that the orientation of the control chart is different, using the vertical axis to display time and the horizontal axis to display the distribution of control results. The IQC design is selected from a drop-down list in the 'design' column of the data entry form. Results from several control materials are entered and their z-values are displayed on the control chart. The 'start-up' design can be recognized from the control limits drawn for the application of a multi-rule IQC procedure. The 'monitor' design has a single set of control limits that correspond to 2.5s. An out-of-control result is signalled by a horizontal bar about two-thirds the way down on the display, after which the start-up design is again applied to provide high error detection and ensure that the problem has been resolved.

Questions and issues

This last section addresses some specific questions and issues about quality management in laboratories today. The answers are not necessarily definitive, but represent my experience and current judgement.

Is IQC still needed, given all the improvements in measurement performance with new generations of analysers?

There is evidence that manufacturers still have their own quality problems in manufacturing and production, and the laboratory must therefore maintain an independent mechanism for assuring the analytical quality of the test results being reported. Proper planning and selection of IQC procedures will usually lead to simpler rules and lower numbers of control measurements for new analytical systems that perform better than previous generations did.

How can it possibly be practical to individualize IQC designs when the laboratory performs hundreds of tests?

In practice, we find it takes only a handful of IQC procedures to give the range of performance needed to cover most tests. For example, in a paper that documents the selection of IQC procedures for a multi-test chemistry analyser, only four different IQC designs were needed to handle the 18 tests being performed by that analyser.[46] Figure 13 shows the critical size errors imposed on the power curves of the candidate IQC procedures. Note that, for 14 tests, the critical errors are actually off-scale to the right. All those tests can be controlled using 3.5s limits and two control measurements per run. Of the remaining tests, albumin

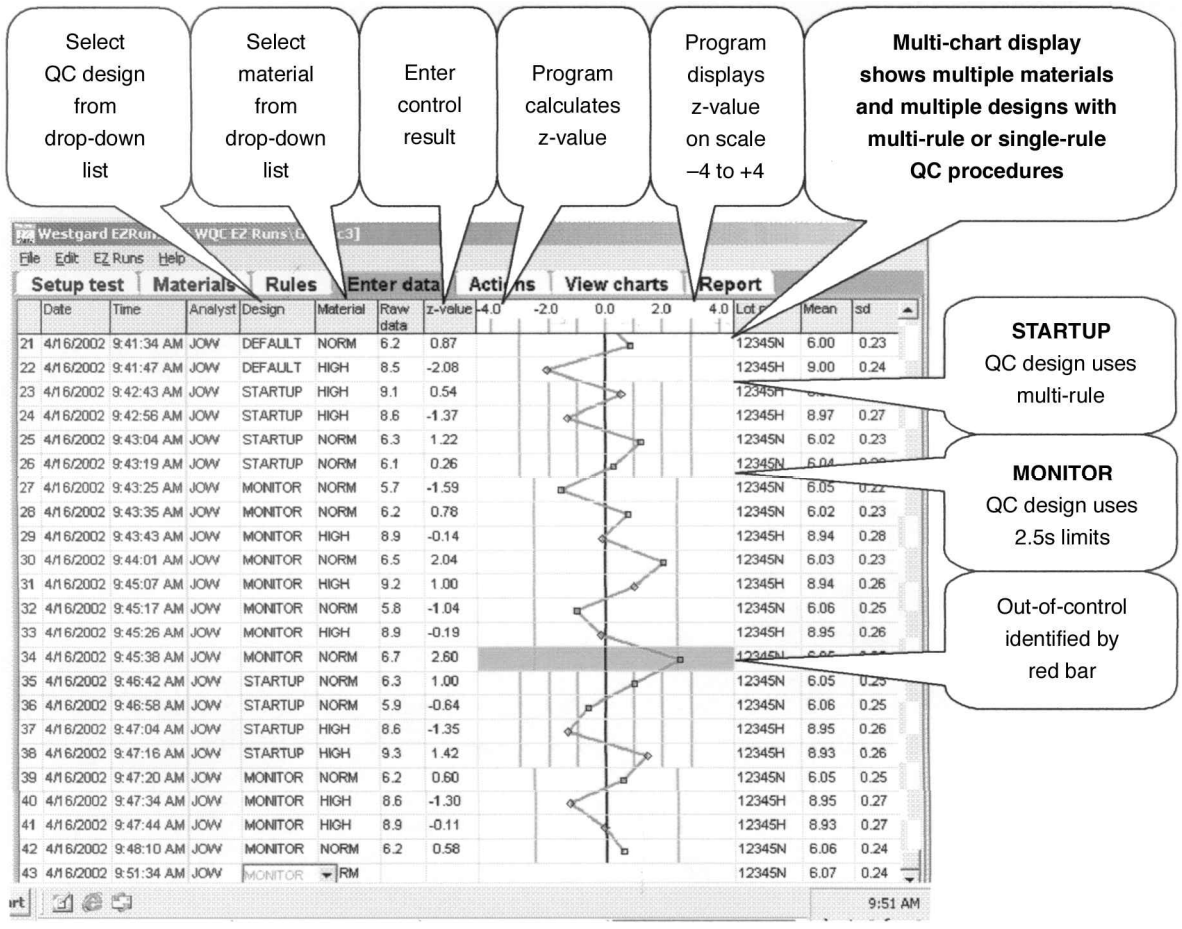

Figure 12. *Implementation of a multi-stage internal quality control (IQC) strategy.*

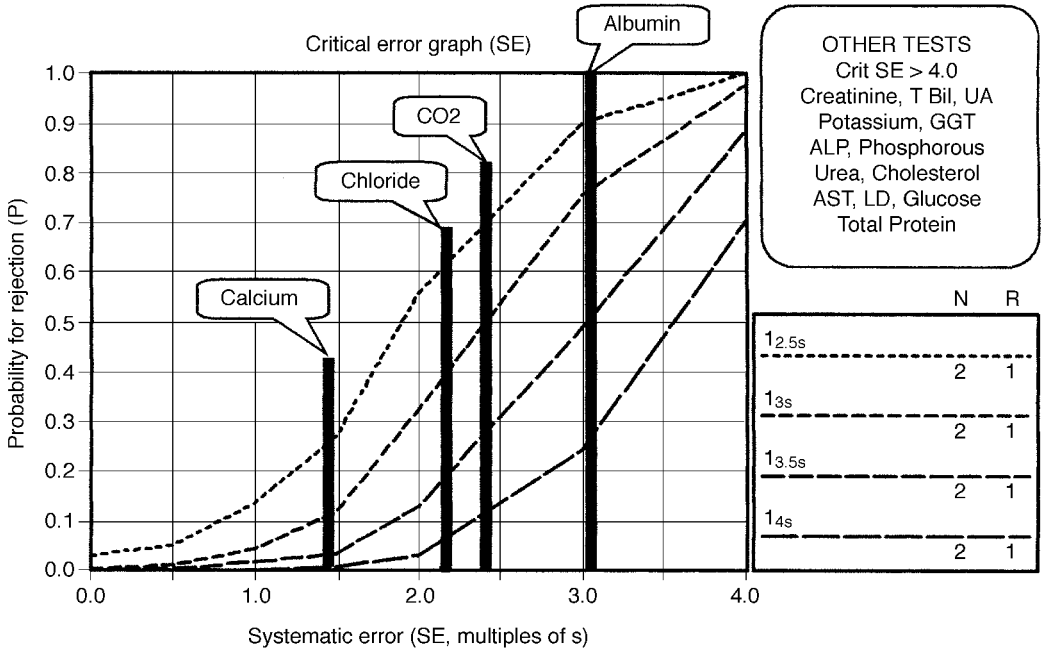

Figure 13. *Results of a sample application of an internal quality control (IQC) design for a multi-test chemistry analyser.* ΔSE_{crit}, *critical systematic error; T Bil, total bilirubin; UA, urine analysis; GGT,* γ*-glutamyltransferase; ALP, alkaline phosphatase; AST, aspartate aminotransferase; LD, lactate dehydrogenase.*

required 2.5*s* limits with an *N* of 2, chloride and CO_2 were candidates for multi-rule procedures and calcium required a special effort that included installing a second test channel to provide duplicate analyses. A sequential study also demonstrated that the laboratory could save money by improving the IQC designs.[47]

How should a laboratory manage the analytical quality of instruments that perform the same tests?

This question reflects the difficulty in matching the performance of different analytical systems, even multiple or identical analysers by the same manufacturer. One approach has been to pool the IQC data from these different systems in an effort to establish an IQC procedure that applies to all the systems. The drawback is, of course, that the control limits no longer reflect the variation observed for the individual systems, and the IQC performance is therefore not well characterized and will vary from system to system. A better approach may be to utilize the same quality requirement, but to account for differences in performance from system to system by selecting the IQC procedure for the individual systems. With this approach, the tests will be managed to the same requirement for quality, but the IQC procedures will not necessarily be the same. All that is required to make this assessment are the OPSpecs charts that represent the quality required for the tests. Several operating points can be plotted to account for the exact performance of each analytical system. If desired, identical IQC procedures can be selected on the basis of the analytical system having the poorest performance.

Isn't it more important today to focus on pre-analytical errors rather than analytical errors?

This argument is often advanced by citing certain studies in the scientific literature that divide laboratory errors into pre-analytical, analytical and post-analytical phases of the total testing process. However, the data in the literature may not be reliable or may not be applicable to laboratory operations today. One frequently cited reference exists only as an abstract,[48] not a peer-reviewed paper, but its credence has grown in significance as the citation is passed from one paper to another. Furthermore, this study predates the US CLIA-88 rules and regulations, reflects practice at a time when laboratories in the USA (and UK as well) were well staffed with well-trained technologists and was carried out in a hospital where the laboratory director was committed to quality.

Can those figures be transferred to healthcare laboratories today?

Are they applicable for near-patient testing (NPT) where personnel have little training and no interest in IQC? While there are more recent studies that are better defined and peer reviewed, such as the one by Plebani and Carraro in Italy,[49] it is still necessary to assess whether the conditions under which the study was performed are representative of the operating conditions in most laboratories today. My opinion is that the results, even of good studies, may not apply to real laboratory operations today because of major differences in the level of staffing, experience and expertise of the staff, and the interest and commitment to quality management, both by the laboratory and by the hospital.

What about electronic QC versus liquid IQC?

IQC is a major difficulty and limitation in implementing NPT, otherwise known as point-of-care testing, and there is therefore a strong interest in doing some other form of QC rather than actually running control samples. Electronic QC fits the need for a simple procedure, but does not usually monitor the analytical factors that may affect the quality of the test result. Electronic QC is still useful for gross checking, e.g. when an NPT device has fallen on the floor, but is not sufficient by itself.[50]

There is clearly a need for new IQC technology for NPT applications. One interesting example is the iQM (intelligent quality management) technology that has been introduced by Instrumentation Laboratory for their GEM analysers. Internal process control solutions are measured very frequently to monitor system performance and to trigger corrective actions that are performed by the system itself. Results of a validation study indicate that the error detection capabilities far exceed that expected of traditional IQC procedures.[28]

Isn't participation in an EQA programme a higher priority than making changes in current IQC procedures?

EQA programmes are well established and provide information that is very useful to the individual laboratory, including the information about method bias. The optimal utility of that information would be obtained by feeding the bias into the IQC planning and selection process to determine the control rules and numbers of control measurements that are needed in that laboratory. Participation in EQA is very desirable, but IQC is essential.

The best approach would be to fully integrate the information from EQA into IQC, as illustrated by Fig. 14. With this integration of IQC and EQA, laboratory data could be automatically processed for purposes of optimizing the IQC procedures being used by the

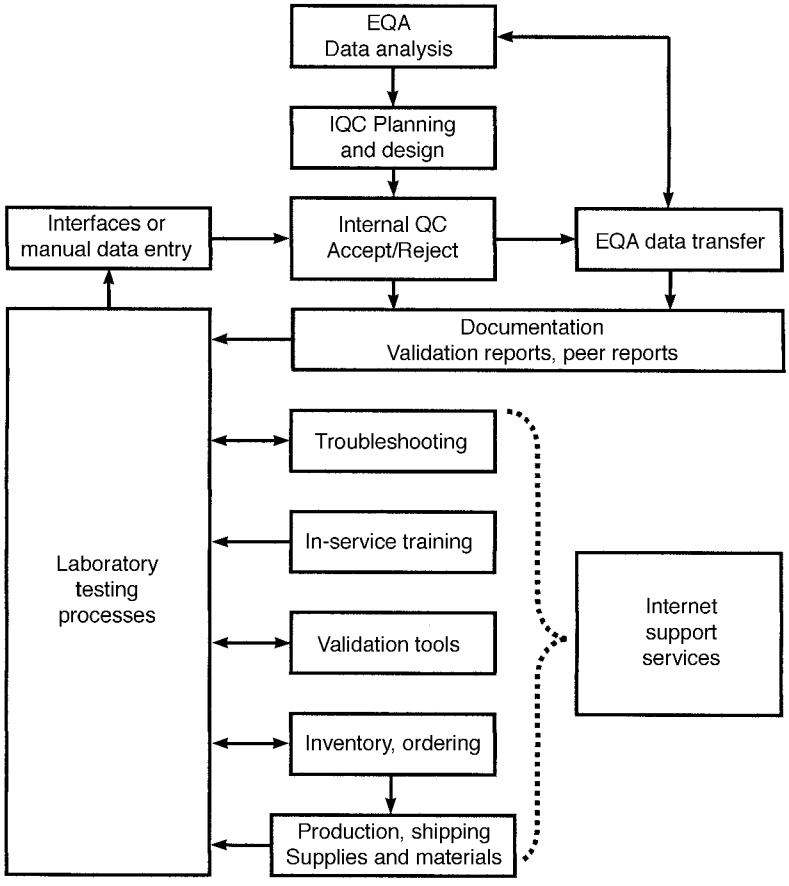

Figure 14. *A total quality control (TQC) support system that integrates internal (IQC) and external (EQA) quality control.*

laboratory. With access to the Internet to speed up data transfer, other Internet services could also be provided to support troubleshooting, method validation, training and the ordering of materials and supplies.

Doesn't the use of clinical or medical validation protocols and programs eliminate the need for improvements in IQC?

This point of view seems to be strongly held in Europe, where computer programs for medical validation are available, such as VALAB[51] and LabRespond.[52] I have no personal experience with these programs, but will still raise the general concern that their performance is not easily documented in a manner that permits comparison with IQC. It is difficult to determine their probabilities for detecting medically important errors, which in turn makes it difficult to compare their performance directly with IQC procedures. In the validation approach recommended by Oosterhuis *et al.*,[52] technical validation via IQC is included as an earlier step in the overall validation procedure. IQC and medical validation should therefore be complementary components of the system for validation of test results. Both are needed, and optimal IQC contributes to a better system for validation of test results.

Nothing but the truth about IQC

It is commonly thought that laboratory tests provide between two-thirds and three-quarters of the information used for making medical decisions. If so, test results had better tell the truth about what's happening with the patients being diagnosed and treated. A test should tell 'the truth, the whole truth and nothing but the truth'. These three dimensions apply to the medical usefulness of a diagnostic test as well as to the analytical reliability of the measurement process.

- Truth requires that a test be related to the disease process of interest and that the interpretative guidelines be understood in terms of the quality (or limits of variation) required for the test.
- For the whole truth to be known, the test must be measured by a reliable method having proper specifications for precision and accuracy.
- To provide nothing but the truth, a test result should not be confounded by unknown or undisclosed factors, such as changes in the subject due to biological variation or changes in the method due to lack of stability or the lack of IQC procedures that will detect those changes.

Glossary of terms

Allowable total error, TE_a An analytical quality requirement that sets a limit for both the imprecision and inaccuracy that are tolerable in a single measurement or single test result.

Analytical run Defined by NCCLS document C24-A2 as an interval, period of time, or series of measurements, within which the accuracy and precision of the measuring system are expected to be stable. In laboratory operations, control samples are analysed during each analytical run to evaluate method performance; therefore the analytical run defines the interval between evaluations of control results. Between quality control evaluations, events may occur causing the measurement process to be susceptible to variations that are important to detect.

Average run length, ARL The average number of runs that occur before a rejection signal is observed.

Average run length for acceptable quality, ARL_{accept} The average number of runs that occur before a rejection signal is observed when there are no errors occurring, except for the inherent imprecision of the method. The ARL_{accept} should be as long as possible and ideally greater than 100.

Average run length for rejectable quality, ARL_{reject} The average number of runs that occur before a rejection signal is observed when there are errors in addition to the inherent imprecision of the method. The ARL_{reject} should be as short as possible and ideally 1.

Control rule A decision criterion for interpreting control data and making a judgement on the control status of an analytical run. It is symbolized here by A_L, where A is the abbreviation for a particular statistic or states the number of control measurement and L is the control limit. An analytical run is rejected when the control measurements fulfil the stated conditions, i.e. when a certain statistic or certain number of control measurements exceed the specified control limits.

Critical-error graph A power function graph that shows the probability of rejection on the y-axis versus the size of errors on the x-axis, with the medically important or critical size error drawn as a vertical line.

Number of control measurements, *N* Used here to indicate the total number of control measurements available for assessing the quality of an analytical run. These measurements may be replicates on one level or material, individual measurements on two or more materials, or replicate measurements on two or more materials. For example, if you assay a single material and make two measurements on that material, N is 2. If you assay two materials and make single measurements on each, N is 2. If you assay two materials and make duplicate measurements on each, N is 4.

Operating specifications Used here to describe the imprecision and inaccuracy that are allowable for a method and the IQC that is necessary to assure, at a stated level, that a defined quality requirement will be achieved in routine operation.

OPSpecs chart A graphical tool that shows the inaccuracy (on the y-axis) and imprecision (on the x-axis) that are allowable for different IQC procedures having a stated level of error detection for assuring a defined quality requirement. The performance of an individual method is imposed by plotting an operating point whose y-coordinate represents the method bias and x-coordinate represents the method imprecision. Appropriate IQC procedures are those whose operational limits (allowable bias and allowable imprecision) are above or to the right of the operating point of the method.

Power function graph A graphical presentation of the performance characteristics of an IQC procedure by describing the probability for rejection (on the y-axis) versus the size of analytical error (on the x-axis) for stated control rules and numbers of control measurements.

Probability, *P* The likelihood an event will occur, usually stated as a decimal fraction between 0 and 1, 0 meaning that the event will never occur and 1 meaning that the event will always occur. For example, $P = 0.05$ means there is a 5% chance that an event will occur.

Probability for error detection, P_{ed} A performance characteristic of an IQC procedure that describes how often an analytical run will be rejected when test results contain errors in addition to the inherent imprecision of the measurement procedure. Ideally, P_{ed} should be 1.00 for errors that are medically important. In practice, a value of 0.90 is often used for selecting and designing IQC procedures.

Probability for false rejection, P_{fr} A performance characteristic of an IQC procedure that describes how often an analytical run will be rejected when there are no errors occurring, except for the inherent imprecision of the measurement procedure. Ideally, P_{fr} should be 0.00. In practice, values less than 0.05 or 0.01 are used for selecting and designing IQC procedures.

References

1 Shewhart WA. *Economic Control of Quality of Manufactured Products*. New York: Van Nostrand, 1931

2 Levey S, Jennings ER. The use of control charts in the clinical laboratory. *Am J Clin Pathol* 1950; **20**: 1059–66

3 Westgard JO. Error budgets for quality management: practical tools for planning and assuring the quality of laboratory testing processes. *Clin Lead Manage Rev* 1996; **9**: 377–403

4 Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1973; **20**: 825–33

5 Westgard JO. Points of care in using statistics in method comparison studies. *Clin Chem* 1998; **44**: 2240–2
6 deVerdier CH, Groth T, Westgard JO. What is the quality of quality control procedures? *Scand J Clin Lab Invest* 1981; **41**: 1–14
7 Whitehead TP. *Quality Control in Clinical Chemistry*. New York: Wiley, 1977
8 Westgard JO, Groth T, Aronsson T, Falk H, deVerdier CH. Performance characteristics for internal quality control: probabilities for false rejection and error detection. *Clin Chem* 1977; **23**: 1857–67
9 Wernimont G. Statistical control of the measurement process. In: *Validation of the Measurement Process*, ACS Symposium Series 63. Washington, DC: ACS, 1977
10 Henry RJ, Segalove M. The running of standards in clinical chemistry and the use of the control chart. *J Clin Pathol* 1952; **5**: 305–11
11 Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clin Chem* 1981; **27**: 493–501
12 Westgard JO, Groth T, Aronsson T, deVerdier CH. Combined Shewhart–cusum control chart for improved quality control in clinical chemistry. *Clin Chem* 1977; **23**: 1881–7
13 Woodward RH, Goldsmith PL. *Cumulative Sum Techniques*, ICI Monograph No. 3. Edinburgh: Oliver and Boyd, 1964
14 Rowlands RJ, Wilson DW, Nix ABJ, Kemp KW, Griffiths K. Advantages of cusum techniques for quality control in clinical chemistry. *Clin Chim Acta* 1980; **108**: 393–7
15 Hoffman RG, Waid ME. The 'average of normals' method of quality control. *Am J Clin Pathol* 1965; **43**: 134–41
16 Cembrowski GS, Chandler EP, Westgard JO. Assessment of 'average of normal' quality control procedures and guidelines for implementation. *Am J Clin Pathol* 1984; **81**: 492–9
17 NCCLS C24-A2. *Statistical Quality Control for Quantitative Measurements: Principles and Definitions: Approved Guidelines*, 2nd edn. Wayne, PA: NCCLS, 1999
18 Petersen PH, Fraser CG, Kallner A, Kenny D. Strategies to set global analytical quality specifications in laboratory medicine. *Scand J Clin Lab Invest* 1999; **57**
19 Westgard JO, Hyltoft Petersen P, Wiebe DA. Laboratory process specifications for assuring quality in the US National Cholesterol Education Program. *Clin Chem* 1991; **37**: 656–61
20 Fraser CG, Hyltoft Petersen P, Ricos C, Haeckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Clin Biochem* 1992; **30**: 311–7
21 Klee GG. Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay specificity. *Clin Chem* 1993; **39**: 1514–8
22 Westgard JO, Wiebe DA. Cholesterol operational process specifications for assuring the quality required by CLIA proficiency testing. *Clin Chem* 1991; **37**: 1938–44
23 Ricos C, Alvarez V. Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, *et al*. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999; **59**: 491–500
24 Westgard JO, Barry PL. Characterizing the quality of statistical control procedures. In: *Cost-Effective Quality Control: Managing the Quality and Productivity of Analytical Processes*. Washington, DC: AACC Press, 1986: 65–91
25 Westgard JO, Groth T. Power functions for statistical quality control rules. *Clin Chem* 1979; **25**: 863–9
26 Parvin CA. Quality-control (QC) performance measures and the QC planning process. *Clin Chem* 1997; **43**: 602–7
27 Parvin CA, Gronowski AM. Effect of analytical run length on quality-control (QC) performance and the QC planning process. *Clin Chem* 1997; **43**: 2149–54
28 Westgard JO, Fallon KD, Mansouri S. Validation of iQM active process control technology. *Point Care* 2003; **2**: 1–7
29 Linnet K. Choosing quality-control systems to detect maximum clinically allowable analytical errors. *Clin Chem* 1989; **35**: 284–8
30 Westgard JO. *Six Sigma Quality Design and Control: Desirable Precision and Requisite QC for Laboratory Measurement Processes*. Madison, WI: Westgard QC, Inc., 2001, 296 pp.
31 Groth T, Falk H, Westgard JO. An interactive computer simulation program for the design of statistical control procedures in clinical chemistry. *Comput Programs Biomed* 1981; **13**: 73–86
32 Westgard JO. Simulation and modeling for optimizing quality control and improving analytical quality assurance. *Clin Chem* 1992; **38**: 175–8
33 Westgard JO. Assuring analytical quality through process planning and quality control. *Arch Pathol Lab Med* 1992; **116**: 765–9
34 Westgard JO. Charts of operational process specifications ('OPSpecs Charts') for assessing the precision, accuracy, and quality control needed to satisfy proficiency testing performance criteria. *Clin Chem* 1992; **38**: 1226–33
35 Westgard JO. *OPSpecs Manual*, expanded edn. Madison, WI: Westgard QC, Inc., 1996
36 Westgard JO. *Basic Planning for Quality*. Madison, WI: Westgard QC, Inc., 2000
37 Westgard Web Tool: Normalized Operating Point Calculator (www.westgard.com/normcalc.htm)
38 Westgard JO, Stein B, Westgard SA, Kennedy R. QC Validator 2.0: a computer program for automatic selection of statistical QC procedures for application in healthcare laboratories. *Comput Methods Programs Biomed* 1997; **53**: 175–86
39 Westgard JO, Stein B. Automatic selection of statistical quality-control procedures to assure meeting clinical or analytical quality requirements. *Clin Chem* 1997; **43**: 400–3
40 Westgard JO, Smith FA, Mountain PJ, Boss S. Design and assessment of average of normals (AON) patient data algorithms to maximize run lengths for automatic process control. *Clin Chem* 1996; **42**: 1683–8
41 Westgard JO, Falk H, Groth T. Influence of a between-run component of variation, choice of control limits, and shape of error distribution on the performance characteristics of rules for internal quality control. *Clin Chem* 1979; **25**: 863–9
42 Neubauer A, Wolter C, Falkner C, Newmeier D. Optimizing the frequency and number of controls for automatic multichannel analyzers. *Clin Chem* 1998; **44**: 1014–23
43 Westgard JO, Groth T, deVerdier CH. Principles for developing improved quality control procedures. *Scand J Clin Lab Invest* 1984; **44 (Suppl 172)**: 19–41
44 Eggert AA, Westgard JO, Barry PL, Emmerich KA. Implementation of a multirule, multistage quality control program in a clinical laboratory computer system. *J Med Syst* 1987; **11**: 391–411
45 Westgard QC, Inc. *Validator EZ Runs Program Manual*. Madison, WI: Westgard QC, Inc., 2002
46 Koch DD, Oryall JJ, Quam EF, Deldbruegge DH, Dowd DE, Barry PL, *et al*. Selection of medically useful quality-control procedures for individual tests done in a multitest analytical system. *Clin Chem* 1990; **36**: 230–3
47 Westgard JO, Oryall JJ, Koch DD. Predicting effects of quality-control practices on the cost-effective operation of a stable multitest analytical system. *Clin Chem* 1990; **36**: 1760–4

48 Ross JW, Boone DJ. Assessing the effect of mistakes in the total testing process on the quality of patient care. *Proceedings 1989 Institute on Critical Issues in Health Laboratory Practice*. Atlanta, GA: Center for Disease Control, 1991: 173
49 Plebani M, Carraro P. Mistakes in a stat laboratory: types and frequency. *Clin Chem* 1997; **43**: 1348–51
50 Westgard JO. Electronic QC, the total testing process, and the total quality control system. *Clin Chim Acta* 2001; **307**: 45–8
51 Valdiguie PM, Rogari E, Philippe H. VALAB: expert system for validation of biochemical data. *Clin Chem* 1992; **38**: 83–7
52 Oosterhuis WP, Ulenkate HJLM, Goldschmidt HMJ. Evaluation of LabRespond, a new automated validation system for clinical test results. *Clin Chem* 2000; **46**: 1811–7

Accepted for publication 18 June 2003