

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/22805>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Proposed Guidelines for the Internal Quality Control of Analytical Results in the Medical Laboratory

Discussion paper from the members of the External Quality Assessment (EQA) Working Group A¹) on analytical quality goals in laboratory medicine

Per Hyltoft Petersen¹, Carmen Ricós², Dietmar Stöckl³, Jean Claude Libeer⁴, Henk Baadenhuijsen⁵, Callum Fraser⁶ and Linda Thienpont⁷

¹ Klinisk Biokemisk Afdeling, Odense Universitetshospital, Odense C, Denmark

² Servicio de Bioquímica, Hospital General Universitario Vall d'Hebron, Barcelona, Spain

³ Institut für Standardisierung und Dokumentation im Medizinischen Laboratorium e. V. (INSTAND e. V.) Düsseldorf, Germany²)

⁴ Instituut voor Hygiëne en Epidemiologie, Klinische Biologie, Brussel, Belgium

⁵ Academisch Ziekenhuis Sint Radboud, Nijmegen, The Netherlands

⁶ Directorate of Biomedical Medicine, Ninewells Hospital and Medical School, Dundee, Scotland

⁷ Laboratorium voor Analytische Chemie, Faculteit Farmaceutische Wetenschappen Universiteit Gent, Gent, Belgium

Summary: The factors involved in analytical quality relate to *definition of quality*, *creation of quality*, and *control of quality*, and errors arise from external and internal sources as well as from permanent and variable factors. Further, the two main types of error are classified as systematic and random errors.

Internal quality control (IQC) systems can only operate on the variable factors which are related to batch-to-batch variations (external factors) and to the performance in the laboratory (internal factors).

In creating an adequate internal control system, several problems are faced:

- (i) quality of control materials,
- (ii) types and frequency of possible errors,
- (iii) number and types of control materials,
- (iv) number of replicates of the control,
- (v) probability of error detection,
- (vi) probability of false rejection,
- (vii) consequences of reject signals,
- (viii) trouble-shooting systems, and
- (ix) prevention of errors among many other conditions.

Gaussian distributions of control results are assumed and the statistical control rules are evaluated in relation to probability of false rejections, P_{fr} , and probability of error detection, P_{ed} , for the different rules. Combinations of low P_{fr} and high P_{ed} are obtained by combining results from e. g. four measurements of the same control sample by use of mean and range rules.

Further, it is not possible to establish a common control system which can be used for all quantities and analytical procedures; on the contrary, each procedure should have its particular efficient IQC system.

These aspects are discussed and a number of guidelines for statistical control rules and problem related internal quality control are presented.

¹) This paper was prepared by a working group (WG-A) which is one of four groups created under the auspices of the European External Quality Assessment Organizers (European EQA-Organizers). The working groups were initiated by Adam Uldall, Denmark and set up following a meeting of EQA-Organizers and interested individuals in Cracow, Poland, in 1991 at Eurolab '91.

²) Present address: Laboratorium voor Analytische Chemie, Faculteit Farmaceutische Wetenschappen Universiteit Gent, Gent, Belgium

Introduction

Internal quality control (IQC) of analytical processes was introduced in clinical chemistry as *Levey-Jennings* charts (1) based on *Shewhart's* theories (2) for 'economical control of quality of the manufactured products'. The control was based on the *Gaussian* (normal) distribution using the $\bar{x} \pm 2s$ (mean + or - twice the standard deviation) rule for acceptance or rejection of an analytical run. In 1977 *Westgard* and 'the Uppsala-group' (3, 4), however, pointed to the fact that about 5% of results from stable runs (series) would be rejected without logical reasons - except by virtue of the chosen control rule - and the situation would be even worse if several measurements of the control samples were introduced in the same run because the proportion of false rejects would increase rapidly. Therefore, they investigated the different control rules by computer simulations and, in their paper from 1979 (5), the theory was made completely clear through the application of *power functions*. Since then, several papers from this group and others have been published in addition to books (6-8); moreover, what have become known as '*Westgard's* rules' have been introduced in the software of many analytical systems. The reader is further referred to some papers of *Linnet* (9, 10).

It should be mentioned, however, that *control* does not necessarily imply *quality*, since control by itself can only be used in monitoring of the current quality of the process, e. g. by rejection of certain errors - but it cannot improve the analytical quality properly. To improve the quality of the analytical procedure, introduction of better methods and calibrators are needed.

The purpose of this paper is to outline aspects of analytical quality in relation to *internal control of analytical quality*, to assess *analytical control rules*, and to provide guidelines for the optimum use of *internal control systems*.

Elements of Analytical Quality

Three aspects can be delineated as the main factors involved in analytical quality. These are *specification of analytical quality*, *creation of analytical quality*, and *control of analytical quality* (11) and are inter-related as illustrated in figure 1.

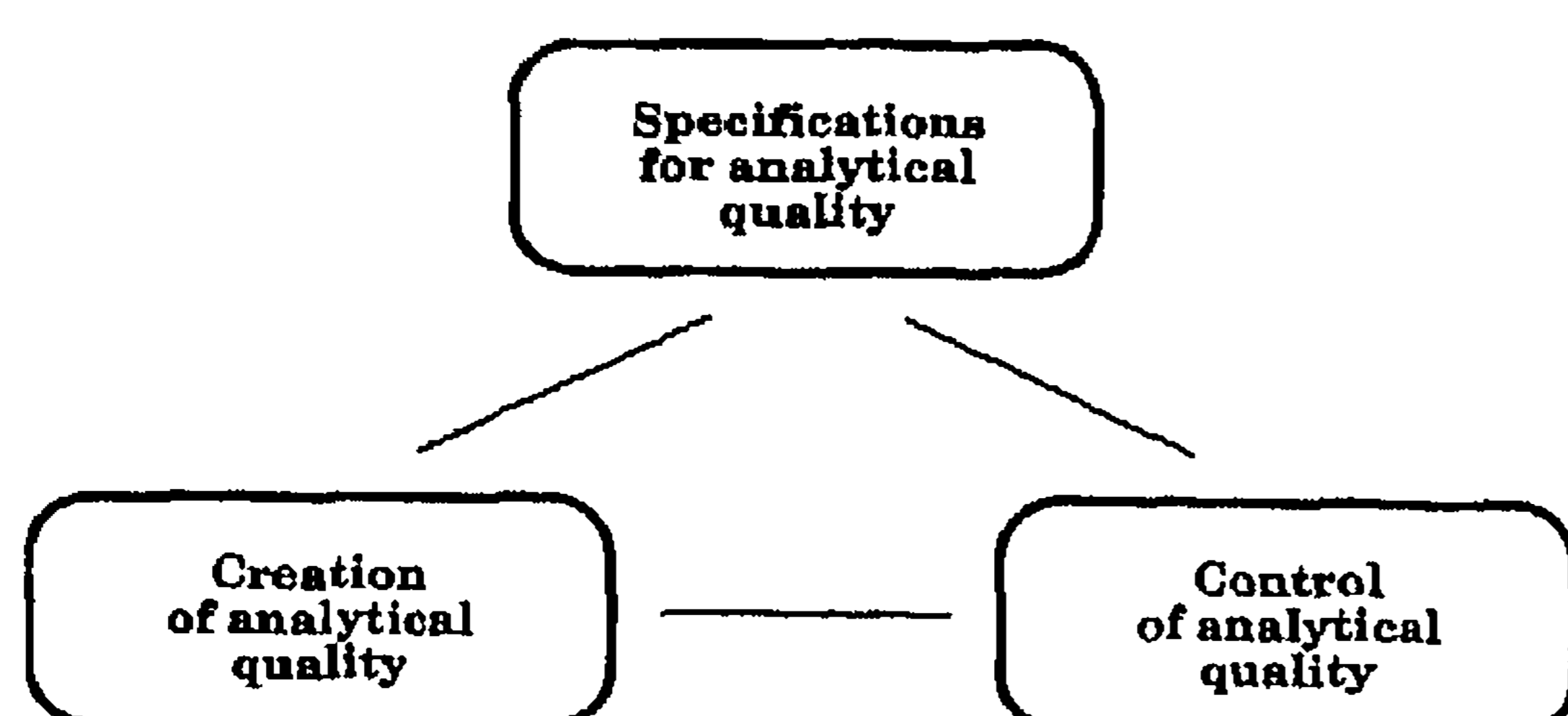


Fig. 1 Elements of analytical quality.

Specification of analytical quality

It is necessary to know what we understand by *analytical quality* in order to judge the quality of a certain analytical procedure. Within laboratory medicine many different approaches to the definition of the quality required have been introduced with a variety of specifications. These have recently been assessed by the Working Group (12) and the biological concept was chosen according to its well documented basis, its simplicity, and its universal application. These recommendations are for maximum allowable analytical imprecision (coefficient of variation), $CV_{\text{analytical}}$:

$$CV_{\text{analytical}} < 0.5 \times CV_{\text{within-subject}}$$

where $CV_{\text{within-subject}}$ is the biological within-subject variation estimated from healthy individuals, and for maximum allowable analytical bias (in %), $Bias_{\text{analytical}}$, when imprecision is negligible:

$$Bias_{\text{analytical}} < 0.25 \times (CV_{\text{within-subject}}^2 + CV_{\text{between-subject}}^2)^{1/2}$$

where $(CV_{\text{within-subject}}^2 + CV_{\text{between-subject}}^2)^{1/2}$ is the combined biological within- and between-subject variation. The recommendation also implies specification for maximum allowable systematic error, ΔSE (to be explained in detail below), when imprecision is negligible:

$$\Delta SE < 0.33 \times CV_{\text{within-subject}}$$

However, when clear and specific clinical strategies for diagnosis and treatment are defined, as for serum cholesterol (13) and for blood HbA_{1c} (14), then analytical quality specifications should be based on the maximum acceptable diagnostic misclassifications. Nevertheless, such strategies are few and the biological concept is, therefore, the best basis for general quality specifications for the majority of the quantities (components) in laboratory medicine.

Creation of analytical quality

The creation of analytical quality is based on external as well as internal factors and on permanent and variable factors as described previously (15, 16) and by the Working Group in relation to external assessment (17); these are illustrated in table 1.

The quality of the external factors are beyond the control of the individual laboratory and, in this regard, each laboratory must try to choose the most appropriate methods (reagents, instruments, calibrators) from those available on the market. This is a difficult task as there is no institution which publishes lists of the quality of commercial products. External assessment schemes which summarize results from the participants ordered according to methods and instruments may be of some help (17), but the concept based on split-sample measurement of patients samples with both reference and routine

Tab. 1 Factors involved in creation of analytical quality

	External	Internal
Permanent	<i>Method:</i> Analytical principle Equipment Reagents (choice of producer) <i>Standardization:</i> Traceability of calibration	<i>Implementation:</i> Choice of conditions 'In house' equipment 'In house' reagents, Time, Temperature, Volume, etc. <i>Standardization:</i> Calibration function
Variable	<i>Batches:</i> Reagents (variability) Calibrators Consumables	<i>Performance:</i> 'In house' reagents Training, Maintenance, etc. Control with 'trouble-shooting' Documentation

methods for evaluation of commercial methods/kits (18) seems to provide a solution to this problem for the quantities where accepted well documented reference methods exist.

The individual laboratory, on the other hand, is directly responsible for its internal factors. First, by correct implementation of the methods, and later through monitoring of analytical quality and detection of errors.

A. Types of analytical errors according to their sources

Before considering internal control systems it is necessary to define the various types of errors. The errors related to the permanent factors (tab. 1) can not be disclosed by any internal control system without help from external quality assessment (EQA) systems (17) or from information of method evaluation groups (18), so the internal control system can only disclose changes in the variable factors. Two types of errors (changes in the variable factors) can be identified, i. e. *external* and *internal errors*, respectively, and these must be treated differently as described below.

A.a. Frequency of internal analytical errors

The frequency of analytical errors cannot be known prospectively but documentation of all errors allows a retrospective estimate to be made, and hence a forecast for the expectation in the near future. This future frequency of errors should, however, be reduced by the following procedure. The types of errors are documented together with frequency of each. For the errors which are easy to prevent, a *fail-safe procedure* is established and documented in the standard operating procedure of the laboratory. Examples are

- a) reagents which can be mixed up (e. g. antibodies for immunoassays) are separated (kept in different freezers) or identified by colour coding,
- b) unstable reagents are used in smaller volumes or not used after a well-defined time;

c) unrealistically small volume pipettings are replaced by larger more reliable volumes;

d) unsatisfactory calibrators (and controls) are changed, etc.

Then the most frequent of the more complicated errors should be tackled. It may be more difficult if the problem is related to expensive components, e. g. the equipment, but more frequent maintenance may reduce the frequency of errors — and replacement should be anticipated in the budget.

A.b. Frequency of variable external errors

As for internal analytical errors it is not possible to know the frequency of external errors — but it is known that they can be expected when a new batch of reagents is taken into use. Therefore, particular attention should be scheduled for every change of lot or batch, in order to reject these if unacceptable, and procure further supplies in good time, i. e., while the old batch is still in use. *Batch comparison* should be made by parallel analysis with the same samples and at the same time. An old batch should not be exhausted before a new one is accessible for comparison.

B. Types of analytical errors according to their effects

Further, errors can be related to *random error* and *systematic error* (19).

B.a. Random error

The random error ("result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions" according to VIM (19)) can be separated into *inherent variation* of the analytical procedure as implemented in the laboratory and *increase in this variation* due to changes in the performance (caused by changes in consumables, reagents, technique, etc.). This change in *imprecision* is quantified by a factor RE

or ΔRE according to *Westgard et al.* (3), where ΔRE indicates the factor by which the *inherent imprecision* is increased.

B.b. Systematic error

The VIM-definition of systematic error ("the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand"; VIM (19)) is not sufficient for laboratory medicine. This is due to the possible sources of error which may be related either to the *calibration* (and some common aspects of performance) which is the same for all measured samples in a run, or to individual (but reproducible) deviations due to *non-specific reactions* or *interfering substances* in the various patient samples (and often different from the control samples). The systematic errors caused by non-specific reactions and interferences (aberrant-sample bias (20)) can be disclosed only by specially designed control systems with selected control materials, e. g. samples from patients with well known pathological composition of interfering substances. This type of evaluation is best performed in external quality assessment (EQA) surveys (21) or by more direct cooperation with producers of reagents (18). In consequence, only 'calibration-type' systematic errors can be dealt with in internal control systems, and this is quantified as ΔSE , where ΔSE denotes the size of systematic error, i. e. the factor in units of standard deviation (for inherent variation) by which the whole run is dislocated.

C. Further characterization of analytical errors

C.a. Persistent errors

Analytical errors can be *persistent*, i. e. be present for longer periods (22, 23). One example is when there are problems with a certain batch of a reagent, where the error will continue until the batch is changed. Persistent errors are relatively easy to discover as they will be repeated day after day and the start can often be related to a change which has been documented in the log-book.

C.b. Intermittent errors

Intermittent errors (22, 23) are difficult to identify as it is often impossible to reproduce the error. Furthermore, this type of error may be unnoticed if the control system has insufficient power.

D. Size of analytical errors

The sizes of analytical errors of course cannot be predicted in general but detailed knowledge of the steps in the analytical procedure may give an estimate. Thus, when two different quantities are measured on one and the same instrument using two different wavelengths, a frequent error might be measurements at the other (and

known) wavelength (when wavelength is not shifted). In such cases, the type and size of the error is predictable. When errors exceed the analytical quality specifications, the run should be rejected, whereas, minor errors should not lead to such consequences. One consequence of minor (in principle – tolerable) errors is, however, a higher frequency of rejection, which in this context may be considered *superfluous rejections* – in order to distinguish them from the so-called false rejections as described below.

E. Other types of errors

When analysts are asked about the most important errors in the laboratory, they will often rank *specimen mix-up* and *incorrect factors applied to diluted samples* highly. Pre-analytical as well as post-analytical errors, however, cannot be disclosed by analytical control systems and they are outside the scope of this document.

Problem-related control

Before going into details of control systems, it is important to note that the control systems described below are not always the best for *error detection* as other checks, like visual inspection of pipettes or check of counts/min (or absorbance) of one of the dilutions of the calibrator, may be more powerful. Furthermore, when the main problems are identified, then *error prevention* may be more important (and often cheaper).

Stable Performance and Errors

Stable performance is characterized by the laboratory's bias [both bias common to all samples due to e. g. calibration (calibrator and calibration function) and the individual bias for each patient (and control) sample due to lack of specificity and to interference in the measurements (matrix effects)] and by random variation, *the inherent imprecision* (determined by the analytical principle, equipment, etc. and the implementation in the laboratory).

Errors are usually defined as the deviation from the *conventional true value*, but in internal control, only deviations from the stable performance can be disclosed. The VIM definition, "deviation" (19), is not specific for these types of problems so the nomenclature outlined by *Westgard et al.* (3, 4) will be retained for the purpose of defining and interpreting internal errors and control.

Errors are here defined as deviations from stable performance. Systematic errors, ΔSE , are deviations, common to all measurements, and random errors, ΔRE , are increases in the inherent imprecision (in principle the Δ is not correct for random errors, but is used currently in the literature).

Control of Analytical Quality

Control can be divided into *internal quality control* and *external quality assessment* according to two separate purposes. Internal quality control can only be used for monitoring the variable factors – and external quality assessment is best for evaluation of the permanent factors as illustrated in figure 2.

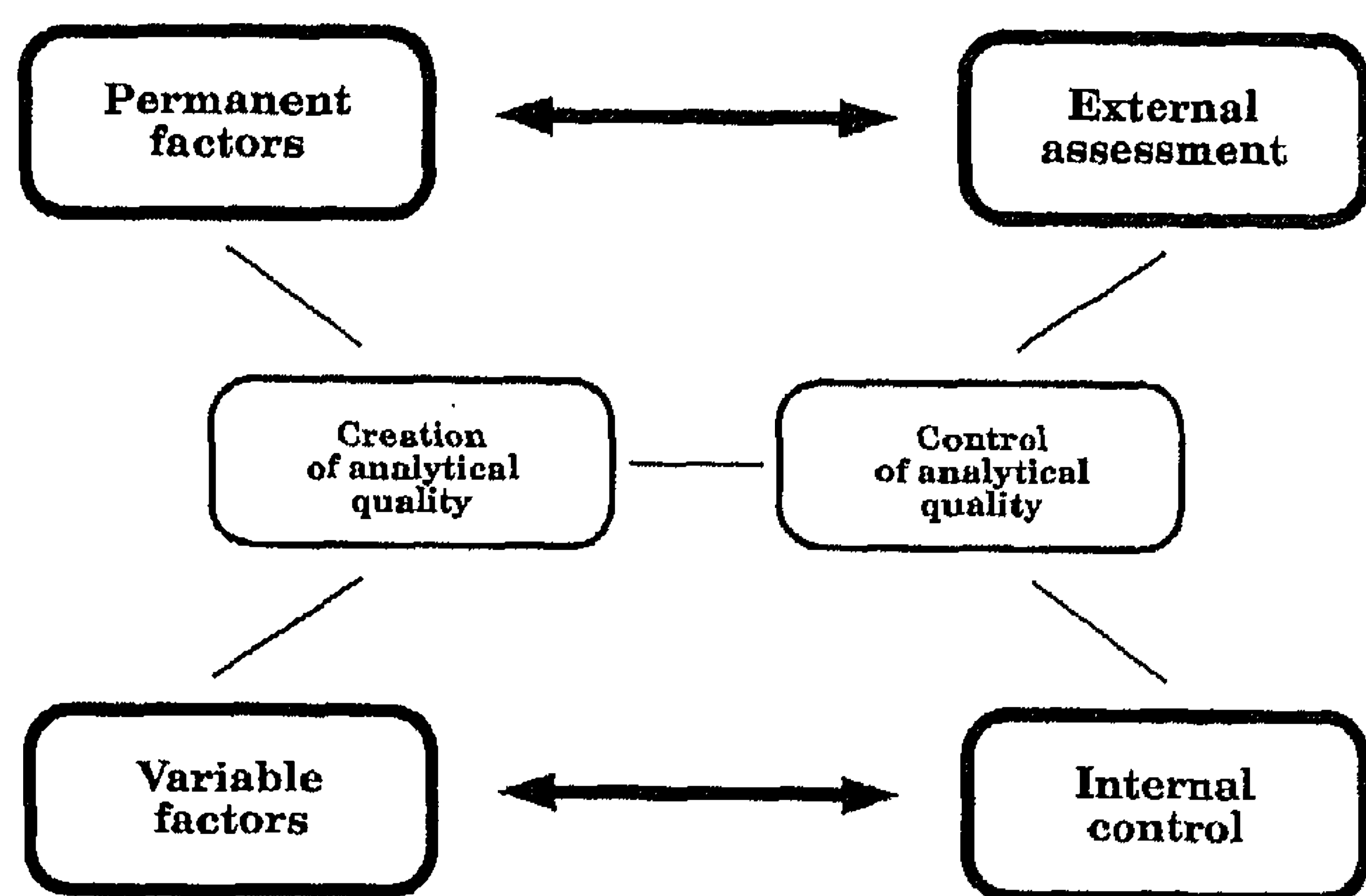


Fig. 2 Illustration of the relations between the factors involved in the creation of analytical quality (permanent and variable) and in the control of analytical quality (external assessment and internal control).

External quality assessment systems may assist evaluation of internal quality control but, in general, external systems are too slow for the ongoing monitoring of performance.

External quality assessment relates to stable performance and it is described in detail in other documents from this Working Group (17, 24).

Internal quality control thus relates to detecting or disclosing deviations from stable performance in the individual laboratory.

Assumptions for Internal Quality Control

In laboratory medicine, samples used for internal quality control are selected samples with values usually known by calculating the mean from several measurements of the sample under 'stable conditions' in the same laboratory. This means that it is an indirect control – in contrast to many industrial products where the performance of the product itself can be controlled directly, e.g., screws and cars, where the characteristics can be measured directly (length, speed etc.). In consequence, the result of the control does not always reflect the quality of the assay of patients samples since lack of precision has different and *random* effect on the samples during the measurements – and will be random again if the measurement is repeated. Further, interfering substances in patient samples may not be reflected by the control samples.

The quality of control materials is of major importance for the quality of the control itself. Therefore, the purpose of the control should be known in order to select the appropriate control material (25, 26). For internal quality control there are two main strategies: *control of the variable external factors* and *control of the stability of the performance*. As described above (fig. 2), the *permanent factors*, such as the traceability of calibrator, analytical principle, and the implementation in the laboratory of the method cannot be controlled by internal control – this process needs an external system. In principle, laboratories can buy (expensive) reference preparations, but this must be considered an external assessment; these are independent of EQA or may be a step in the process of quality improvement, and thereby, outside the scope of internal control. However, in the case that much cheaper commutable IQC materials with reference method target values are available in the future, traceability can also be controlled internally. The dilemma of expensive commutable materials with traceable values might be solved if these materials are used for internal control as well, which would need much greater volumes and thereby – maybe – cheaper materials. Genuine (non-processed) serum with traceable target values is, however, difficult to handle. It must be stored at -80°C and mailed on dry ice, as in the Nordic protein project (11), so further efforts are needed to solve this.

In principle control rules do not differ, irrespective of whether used for detection of the variable external errors or internal errors, but trouble-shooting and possibilities for correcting errors are different for these. In order to detect external variable errors, a painstaking record on each batch is needed and in principle every new batch should be tested before use (cf. ISO 9000-series). A focus should be kept on changes in the external factors and the control results interpreted accordingly. The same should be the case for internal changes but here many other factors are involved so the maintenance of the analytical procedure may be as important as the internal

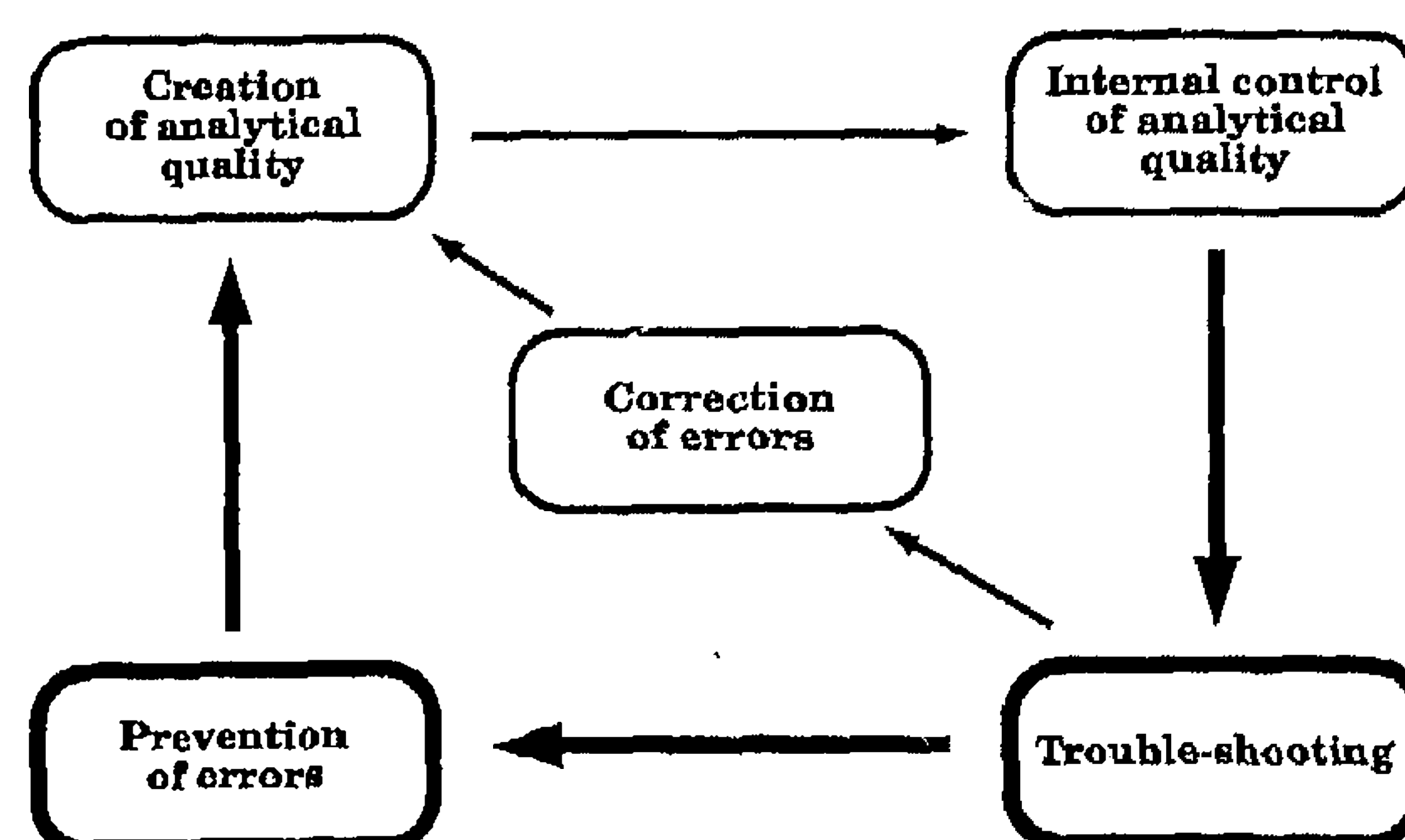


Fig. 3 Model for reaction-pattern for rejection-signals from the internal control system: When a rejection signal occurs, then trouble-shooting should start and the error(s) should be corrected. However, this should lead to establishment of fail-safe procedures in order to prevent these types of errors in the future.

control. The relationships between *internal control, trouble-shooting, correction of error, prevention of errors* (fail-safe procedures), and establishment of *creation of (better) analytical quality* are outlined in figure 3.

It is essential to understand that internal control has its limitations and that it is only a part of the quality of the analytical process – but it is a very important tool for detecting errors in each batch and changes in performance when it is handled correctly and with care and reflection.

Materials for Internal Control

The most important characteristic of any material used for internal control is its ability to reflect deviations from stable analytical performance. Therefore, it must be stable and reproducible, i. e., the measurand must be in exactly the same molecular form and the same concentration each time it is used for control – and the surrounding components (the matrix) must also be identical. These characteristics are indispensable for materials to be used for internal control according to the statistical control rules described below. Control materials, e. g. for serum analyses with composition close to genuine serum, are usually better than artificial materials if they are stable and reproducible. There have been no commercial QC-materials available until now with these quantities, but in many laboratories pools with fresh serum from healthy individuals are used for internal control of serum proteins and a number of hormones. In some countries, the use of single donations of patient serum or serum pools is also allowed. In view of the above mentioned limitations lyophilized material is often preferable to other materials, but the reconstitution of the volume must be painstakingly performed. This is in contrast to external assessment, where genuine control materials are superior to other materials (25, 26). Likewise, 'true' concentration values are not essential for internal control provided that the target values are determined under optimal conditions (stable performance). The assignment of values could be performed while high quality calibrators are used for calibration and/or external assessment with high quality materials is performed. This could take place over, e. g. 20 analytical runs co-ordinated with the yearly external assessment as described by Libeer et al. (17). Internal control is thus linked to the external assessment and both are indispensable for documentation and monitoring of the analytical quality.

Reaction Patterns for IQC Rejection Signals

When reaction patterns for actions taken on reject signals are considered, it is of little help to search the avail-

able literature. Here only the strict interpretations of such signals will be found. The way to investigate this is to listen to discussions among analysts or to ask indirect questions; otherwise the problems will be missed. However, interpretation is often different for various equipment in the same laboratory and it is difficult to be objective about how the patterns were introduced.

Examples of reaction patterns for reject signals

a. The general type

The general type is an immediate rerun of the series. This is also the generally accepted concept for handling reject signals. However, this pattern is as bad as many other reactions (except for a few situations as described below). The reasons are because there are two possibilities:

- a) either the run was satisfactory and it was a false rejection due to a poor control rule which results in the work being unnecessarily doubled, or
- b) there was an error and then it is not expedient to repeat the error and so again the work is doubled unnecessarily.

The relevant solution to the problem is to *introduce control rules with low probability of false rejection* as described below and, when these are established, then the reaction to reject signals should be, to *stop the work – start trouble shooting – correct the error – and, if possible, implement a fail-safe procedure* as illustrated in figure 3. Most will recognize subjective decisions, for example, where the analyst says that a method has been re-run three times and the control is still outside the acceptance limits. Then it might be said (after considerations) that the error is *without clinical importance* and *the report shall be issued*. Of course, clinical relevance, and thereby the size of acceptable error, should have been decided beforehand (based on thorough and careful considerations) when the analytical quality specifications were outlined and the control rules were introduced and not in stressed situations.

It may be a troublesome process to identify the many different errors and find fail-safe procedures – but it is the best way to solve problems and to eliminate the illogical feeling of doing a good job when re-running several times and performing doubtful decisions based on incomplete facts.

Exceptions from the strict pattern of trouble-shooting are systems with multiple channels and procedures where the analytical time for a single measurement is very short and samples are processed separately, e. g. many haematological analyzers and near-patient instruments. Here, it is so easy to repeat the control measure-

ment that it is logical to do this before trouble-shooting is started (see below).

b. The overlook type

The overlook type is the *lack of reaction* to reject-signals. This situation is not intended, but arises under certain conditions, e. g. when the control system has a very high frequency of false rejects or a persistent error has lasted (uncorrected) for a long time. This could be a batch of low quality or a control material with a poorly defined target value. In the latter case, the neglect of reject-signals might be correct, but this is not known.

c. The multi-channel type

The multi-channel type is related to the high frequency of repeated testing (control signals) performed in multi-channel analyzers, when the control rules are not adjusted for each of the many tests. The probabilities of rejections according to the number of repeated testings using this control rule are listed in table 2.

Tab. 2 Probability of false rejection of runs in relation to number of repeated analyses

Number of repeated analyses	Probability of false rejections P_{fr} for $\bar{x} \pm z \times s$	
	$\bar{x} \pm 1.96 s$	$\bar{x} \pm 2 s$
1	5%	4.6%
2	10%	8.9%
5	23%	20.8%
10	40%	37.2%
20	65%	60.6%

$P_{fr}(\text{repeated}) = 1 - [1 - P_{fr}(\text{single})]^n$, where n is the number of repeated analyses and P_{fr} in the formula is measured in fraction.

Some laboratories make recalibrations of all runs by use of the 'controls', whereas others "only" recalibrate when the control is outside the $\bar{x} \pm 2 s$ limits. In both situations the rejection rate is reduced, but at the same time the analysis is without any kind of control system and the quality is unknown. Again the solution is to establish proper control rules with low probability of false rejects.

d. Other types

Several other types of handling control systems wrongly may be possible, but it is more interesting to evaluate the correct control systems.

Theoretical Basis for Statistical Control Rules

The inherent imprecision is assumed *Gaussian* in nature and also the whole theory is based on this assumption. Results from both patient samples and controls are assumed to vary independently according to the inherent imprecision. In principle any deviation, whether system-

atic or random, should be reflected to the same extent in the results for both patient and control samples. Thus, the control result is an indicator of the quality of the run and, hence, of the quality of the patient results. This is not completely true for random errors as they vary independently and it is interesting in the situation where the control has an 'abnormal' result which is actually still a component of the stable performance. With a control rule of $\bar{x} \pm 2 s$ this will happen in about 5% of all runs, resulting in a reject of a run which in principle is acceptable. By using another control rule, e. g. $\bar{x} \pm 3 s$ (mean $\pm 3 s$), the probability of 'false rejects', P_{fr} , will be reduced to 0.3%. Thereby, the unnecessary rejects are reduced considerably, but at the same time the probability of error detection, P_{ed} , is also reduced. The P_{fr} and the P_{ed} are functions of the control rules and, by selection of reliable control rules, the analytical work can be smoothed and also be more reliable. The P_{fr} -values for a number of simple control rules of the type $\bar{x} \pm z \times s$ are listed in table 3.

Lack of understanding of the basic concepts of statistical control rules has led to some illogical procedures as described above. By choosing the relevant control rules for the internal control system, however, it becomes possible to introduce systems with control rules which can keep the P_{fr} low and at the same time allow for a high P_{ed} as described below.

Control rules

Both patient samples and controls are assumed to be distributed randomly (*Gaussian*) during an analytical run. This means that a control result may be outside the control limits — according to the P_{fr} — whereas patient samples are determined correctly, and vice versa. Therefore, the purpose of the design of control rules is to distinguish between false rejections and errors, and often to indicate the type of error (random or systematic); in consequence, trouble-shooting procedures have already started. As long as only one control sample in single determination is used, the possibilities are limited to control rules of the type $\bar{x} \pm z \times s$ — but an understanding of the simple rules is important for more complex concepts. Understanding P_{fr} should be easy from table 3

Tab. 3 Probability of false rejection with a single analysis using control rules of the type $\bar{x} \pm z \times s$

Control rule	Probability of false rejection, P_{fr}
$\bar{x} \pm 2 s$	5% (4.55%)
$\bar{x} \pm 2.575 s$	1%
$\bar{x} \pm 3 s$	0.3%
$\bar{x} \pm 3.5 s$	0.05%
$\bar{x} \pm 4 s$	0.01%

and the interpretation of P_{ed} becomes clear from the power-functions (see below).

Power functions

For a certain control rule and a certain type of error the probability of detecting the error, P_{ed} , can be obtained from statistical tables or estimated from computer simulations.

One control value only

By combining P_{fr} and P_{ed} as a function of the sizes (and types) of errors in a plot it is possible to get a clear picture of the relationships by the resulting *power-functions* as illustrated for systematic errors in figure 4 and for random errors in figure 5, using two different control rules, mean ± 3 or 2 times the standard deviation, respectively.

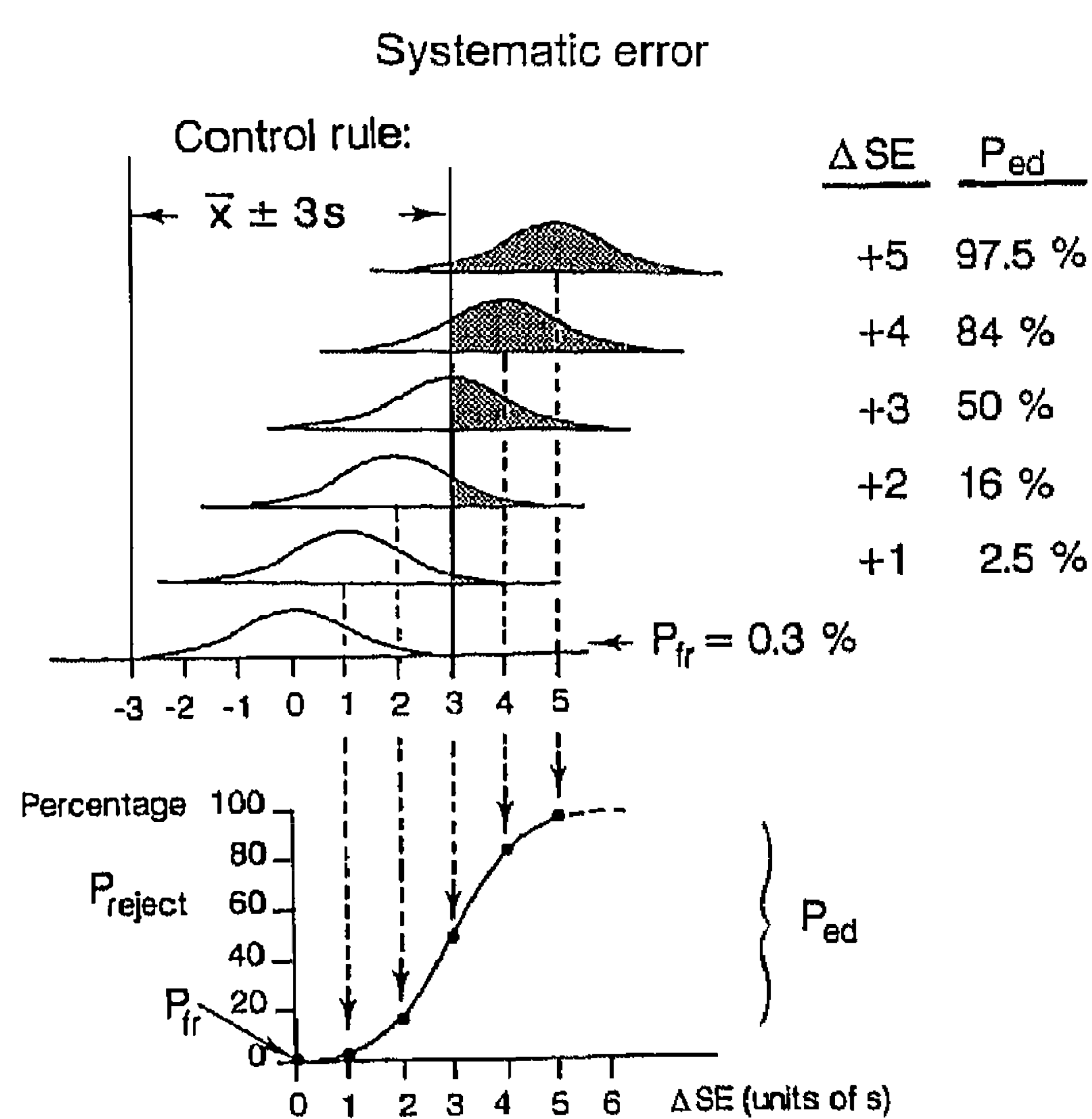


Fig. 4 Illustration of the relationship between systematic errors (upper left) from ΔSE zero to +5 s and (lower) power function for the control rule $\bar{x} \pm 3s$. The P_{ed} -values are shown (upper right). The ordinate, P_{reject} , denotes P_{ed} and the single point of P_{fr} .

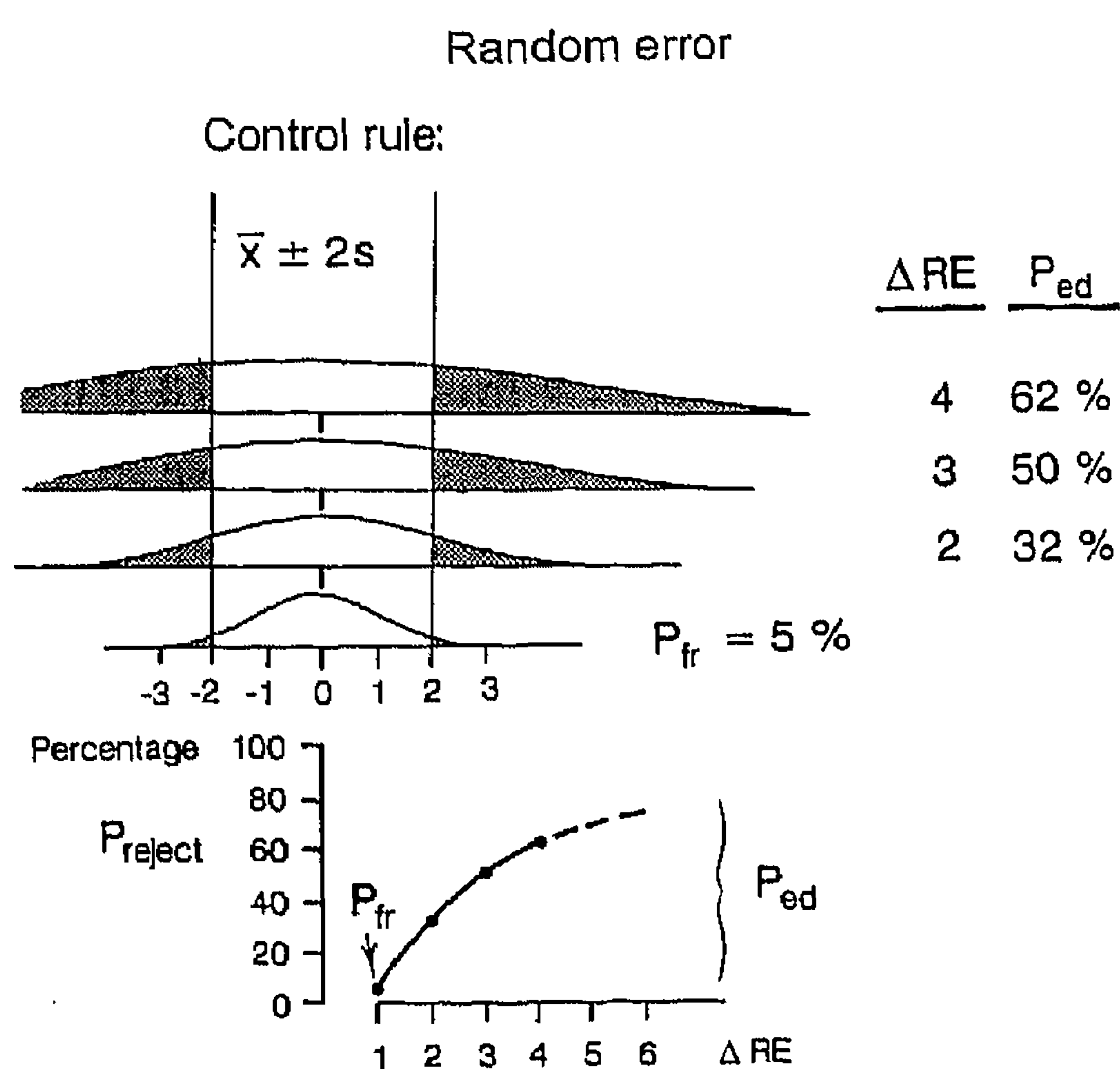


Fig. 5 Illustration of the relationship between random error from $\Delta RE = 1$ (inherent imprecision = s) to 4 and power function for the control rule $\bar{x} \pm 2s$. Composition as in figure 4.

Figure 4 illustrates the effect of increasing systematic errors (all with the same original s). The shadowed areas denote the probability of rejection which is projected to the plot below, the power-function. The extreme situation of no error ($\Delta SE = 0$) is the point of P_{fr} , whereas the rest of the curve describes the continuous P_{ed} -function. When the systematic error exceeds the control limit ($3s$) by more than 2 times the standard deviation, i. e. for $\Delta SE > 5$, then the P_{ed} exceeds 97.5%, which means that a systematic error of this size will rarely be missed by the control system (only one out of 40). The power for detecting negative systematic errors is identical to the positive errors.

Compared to control systems for systematic errors, these simple control rules are less powerful for random errors of considerable sizes as the curve only slowly approaches the 100% line (figure 5). The powers for detecting the two types of errors are thus different but, as only one control rule has been used, the P_{fr} is related to the control rule only (0.3% in fig. 4 and 5% in fig. 5), which means that P_{fr} only depends on the control rule chosen. A series of power functions for systematic errors are shown in figure 6.

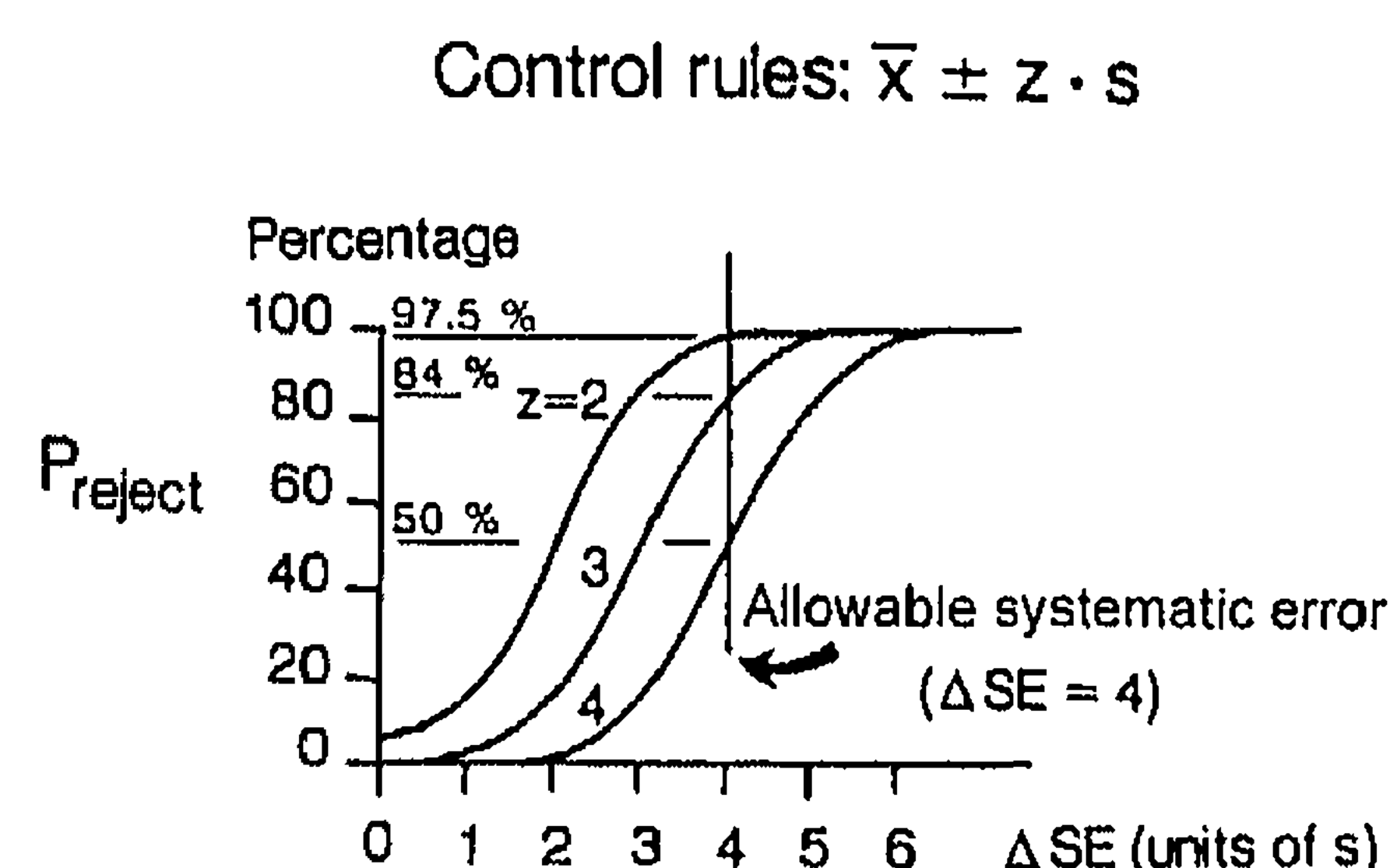


Fig. 6 Power functions for systematic error and for the three control rules $\bar{x} \pm 2s$, $\bar{x} \pm 3s$, and $\bar{x} \pm 4s$. An assumed allowable systematic error of $\Delta SE = 4$ is indicated. See table 3 for the respective P_{fr} -values.

Three control rules of the type $\bar{x} \pm z \times s$ with $z = 2, 3$, and 4, respectively, are illustrated in figure 6 and the *allowable systematic error* (defined by analytical quality specifications) is here set to $\Delta SE = 4$. The three control rules (with P_{fr} equal to 5%, 0.3%, and 0.01%), will have probabilities for detection of the systematic error (P_{ed}) of 97.5%, 84%, and 50%, respectively.

Replicate control measurements in a single run

By performing replicate measurements of the control sample within an analytical run and calculation of the mean of the replicates, \bar{x} the variation of this mean, $s_{\bar{x}}$, is reduced by the square root of the number of replicates. Using four replicates, $s_{\bar{x}} = s/2$, which makes it possible to combine a low P_{fr} with a high P_{ed} as illustrated in figure 7 for the control rule $\bar{x} \pm 4 \times s_{\bar{x}}$. This control rule has the same P_{fr} as the $\bar{x} \pm 4s$ rule and a higher P_{ed} for

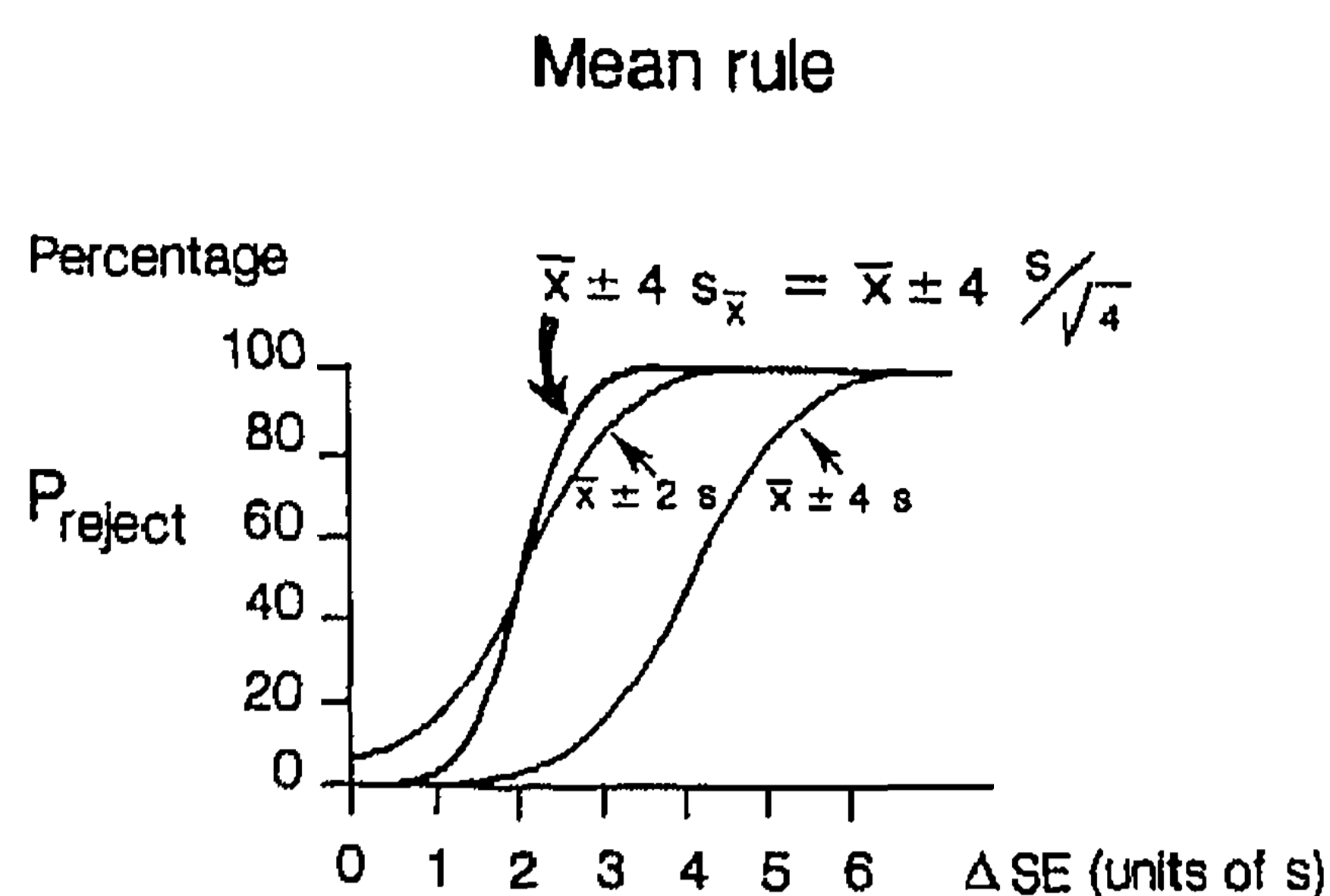


Fig. 7 Illustration of the advantages of the mean-rule by combining a low P_{fr} -value with high power for detecting systematic errors. The mean-rule $\bar{x} \pm 4 s_{\bar{x}}$ (of four replicates), combines the P_{ed} of the $\bar{x} \pm 2 s$ rule and the P_{fr} of the $\bar{x} \pm 4 s$ rule (both with 1 determination).

the systematic error $\Delta SE = 4$ than the $\bar{x} \pm 2 s$ rule (for relations between within- and between-run variations, see l. c. (27)).

This example demonstrates how powerful the combination of several control results can be. *Westgard* and 'the Uppsala-group' have described the power-functions for a long list of different control rules (5) and also with respect to separation of imprecision into within- and between-run variation (27). Moreover, several books on the subject have been published (6-8). It is out of the scope of this article to go into detail of the many possible rules and their power-functions, except for underlining the types of the best control rules for systematic errors: *Mean-rules* and *Cusum-rules* (in the latter results are cumulated over time) and for random error (χ^2) and the *range-rules*. In the 'Validator-programme'

described below³⁾ many of these rules are included, and a few will be described in these guidelines.

Control Rules to Guarantee Quality with a Certain Probability

For the analytical quality specifications recommended by two European groups (12, 28) based on biological criteria and specified for both bias and imprecision, control rules can be decided from power-functions (5, 7, 8, 27) when the inherent imprecision is known, i. e. when imprecision is estimated during stable analytical performance.

The allowable total systematic error, SE_T ($SE_{tolerable}$) (6), is defined by the tolerable sum of the permanent analytical bias, B_A , and systematic error of runs (deviation from B_A), ΔSE :

$$SE_T = B_A + \Delta SE$$

and the allowable ΔSE can be calculated as the difference between SE_T and B_A , which thereby depends on the permanent bias, estimated from external quality assessment or other 'traceable' materials. The power-functions for four different control rules: $\bar{x} \pm 2 s$ with two measurements of the control, $\bar{x} \pm 3 s$ with four replicates, and two control rules of the type $\bar{x} \pm z \times s_{\bar{x}}$, both with four replicates, where $z = 2.58$ and 3, corresponding to $P_{fr} = 1\%$ and 0.2% (and symbolized by $x_{0.01}$ and $x_{0.002}$), respectively, are illustrated in figure 8,

³⁾ QC Validator™, Westgard® Quality Corporation, 112 Shore Road, P. O. Box 2026, Ogunquit, ME 03907, USA.

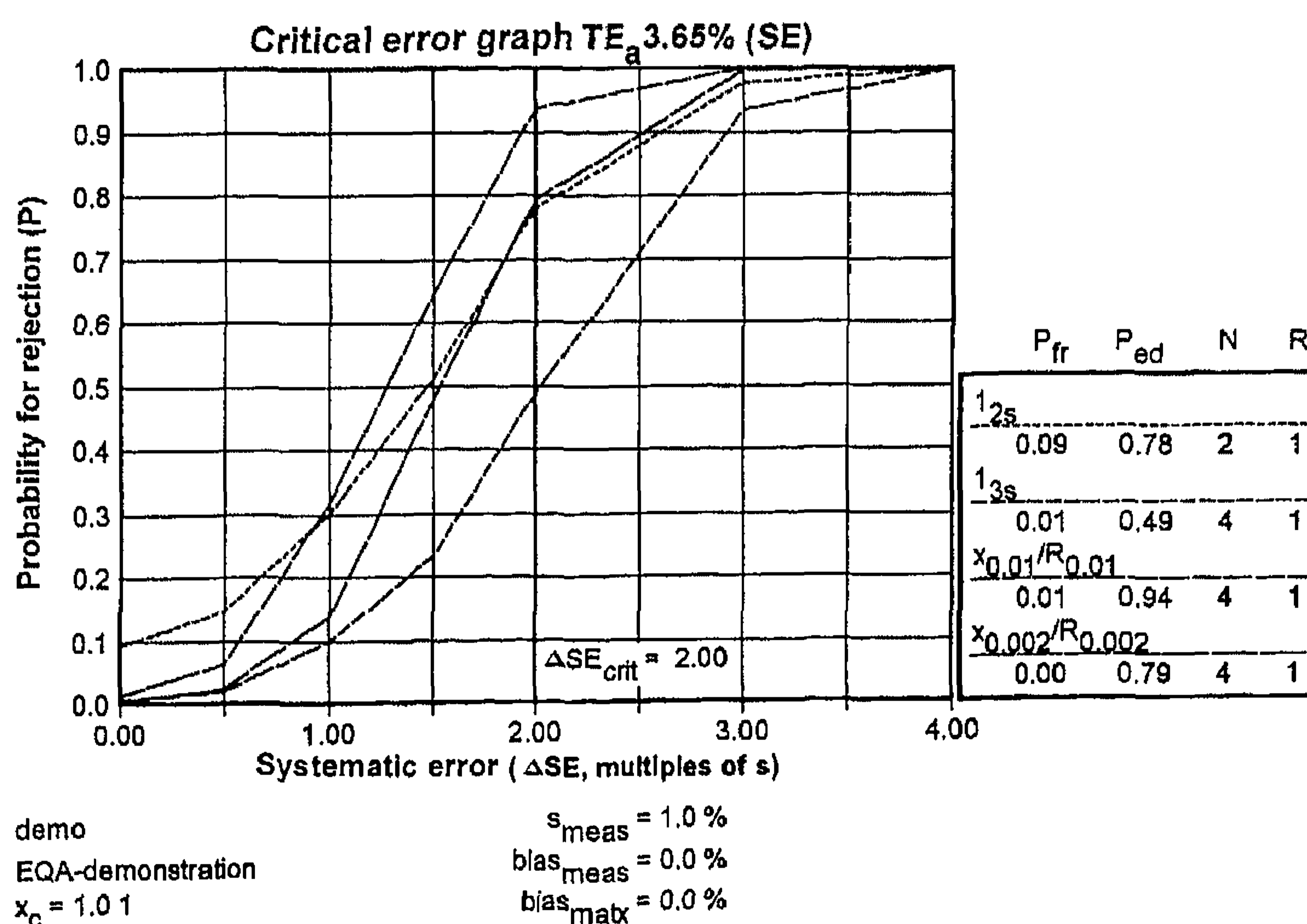


Fig. 8 Power-functions for systematic error for four different control rules $\bar{x} \pm 2 s$ (2 replicates), $\bar{x} \pm 3 s$ (4 replicates), and the mean-rules $\bar{x} \pm 2.58 s_{\bar{x}}$, and $\bar{x} \pm 3 s_{\bar{x}}$ (both 4 replicates). P_{fr} -values of approx. 9%, 1%, 1%, and < 1% are shown together with the P_{ed} -values for $\Delta SE = 2$ in the box at the right. Deviations in P_{fr} - and P_{ed} -values from theoretical are due to the computer simulation

technique. Print out from the QC Validator™-programme (see footnote³⁾) with permission.

- demo = example
- s_{meas} = s measured
- $bias_{meas}$ = bias measured
- $bias_{matr}$ = bias due to matrix

Reduction of inherent imprecision

It is evident that improvement of the inherent precision, e.g. by change of instrument, will solve a lot of the above mentioned problems. First it improves the general analytical quality, but second, a reduction of imprecision to half the value will have the same effect as the mean rule in figure 7 (when calculations are performed using the "old" standard deviation).

The Validator-Programme

Westgard has designed a computer programme for optimizing of control rules according to the above described principles and in relation to analytical quality specifications³). The computer programme presents the current knowledge on internal control rules for laboratory medicine based on the pioneering work (3–5) and on the relations to analytical quality specifications (28, 30–33). The two papers from 1994 compare analytical quality specifications derived in Europe (28) and in USA (34, 35).

Using the programme, it is possible to evaluate specific control rules for an individual laboratory, when analytical imprecision and bias are known. The programme, however, is designed for the US-CLIA-criteria, which are total error criteria, so the interpretation of the figures needs some introduction.

In order to use the programme for European analytical quality specifications the values for allowable bias,

B_{allow} , and allowable imprecision, s_{allow} , the *total allowable error*, TE_{allow} , to be used is calculated as

$$TE_{\text{allow}} = B_{\text{allow}} + 1.65 \times s_{\text{allow}}$$

The factor 1.65 is valid for the 95% tolerance limits (2.33 for 99%).

Short description of the 'Validator programme'

The 'Validator programme' is a computer programme for planning of internal control systems (control rules and number of replicates) when analytical quality specifications are defined and the stable analytical performance is known by its stable analytical bias and inherent analytical imprecision. The programme provides quick access to

- power-function graphs,
- critical-error graphs and
- 'OPSpecs carts' (operational process specifications).

By entering the data for stable analytical imprecision and bias as well as for allowable total analytical error (plus additional information of the analytical quality, concentration for critical interpretations and matrix bias, which are not essential for understanding of the principles) the three options can be investigated by graphs with tables. A great number of control rules with different numbers of replicates can be chosen and the appropriate graphs are presented:

- Power-function graphs are like figures 6 and 7, but not smooth due to the computer simulation technique

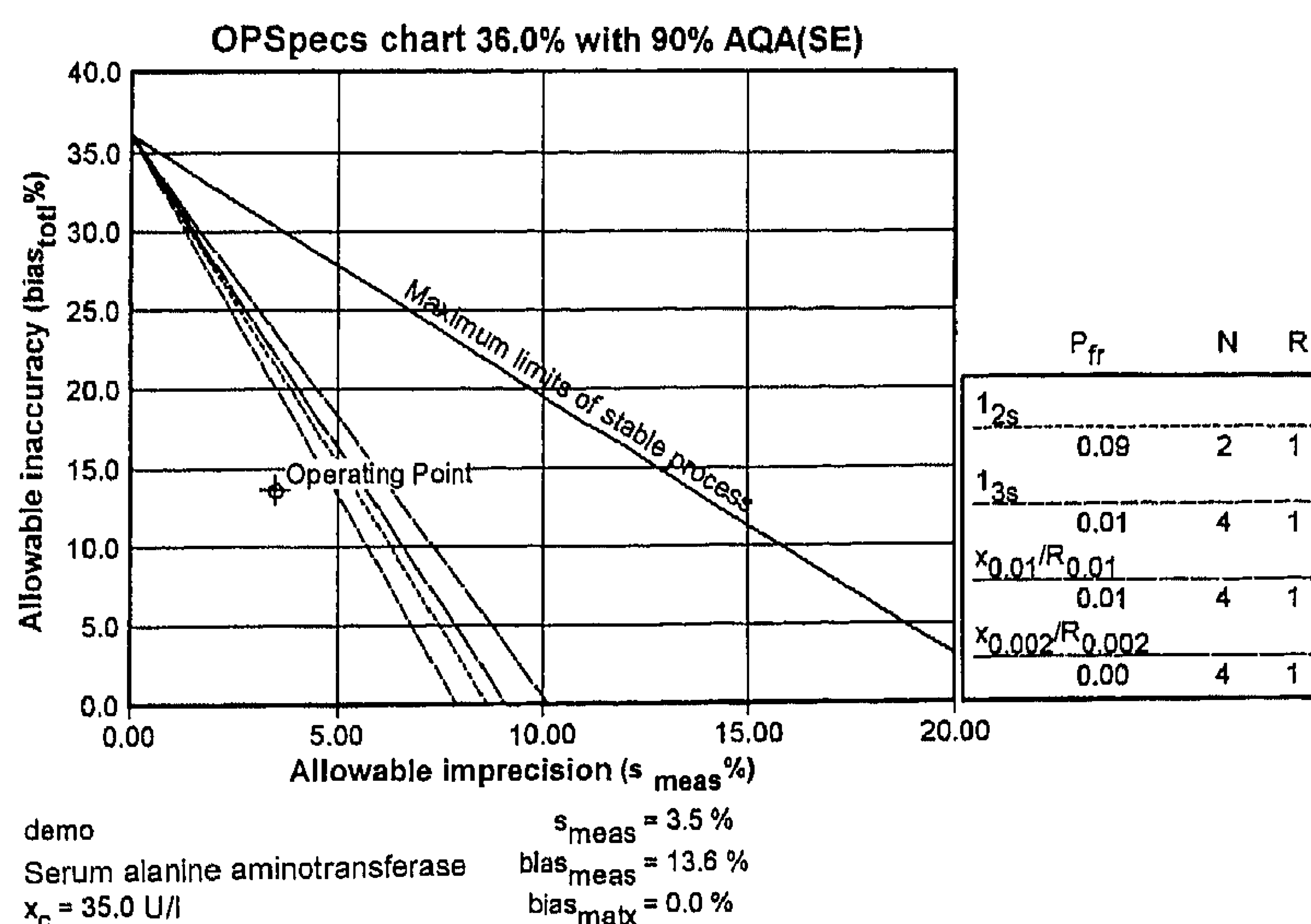


Fig. 10 OPSpecs Chart for serum alanine aminotransferase with 'Allowable Total Error' equal to 36%, from the allowable systematic error 13.6% and allowable random error 13.6% ($13.6 + 1.65 \times 13.6$) illustrating the same control rules as in figures 8 and 9 and with a power of 90%. In this example an analytical bias of 13.6% and an inherent imprecision of 3.5% are assumed, giving the 'Operating Point', which must be 'within' a control rule (to the left) if the power is sufficient. Deviations in P_{fr} - and P_{ed} -values from theory are due to the computer simulation technique. The

heading is short for "operating specifications chart for 36% total error with 90% achieved quality assurance for systematic error". Print out from the QC ValidatorTM-programme (see footnote³) with permission.

demo = example
 s_{meas} = s measured
 $\text{bias}_{\text{meas}}$ = bias measured
 $\text{bias}_{\text{matr}}$ = bias due to matrix

and lack of curve-fitting as seen from figures 8 and 9 (which include critical errors).

b) Critical error graphs are identical to the power function graphs, but with additional introduction of the 'critical error' (figs. 8 and 9).

c) OPSpecs charts are more complicated (figs. 10 and 11). The plot illustrates for different control rules the allowable total systematic error as function of allowable imprecision. Each line demonstrates the functional relationship (for the chosen control rule and number of replicates) between the two. A large systematic error is related to a small random error, and vice versa. The line to the right illustrates the limits for the hypothetical situation of total stability and without control rules. The 'op-

erating point' indicates the combination of the actual (and known) analytical bias and imprecision (inherent imprecision). This point must be to the left of the control line in order to assure the power (probability of error detection) as indicated in the top of the figure.

The programme also has an option for combining the analytical imprecision with the within-subject biological variation. This option, however, is more complicated and will not be referred to here.

Examples on design of control rules

1. According to the European recommendations (28) the allowable bias and imprecision for the quantity serum alanine aminotransferase are 13.6% and 13.6%, respec-

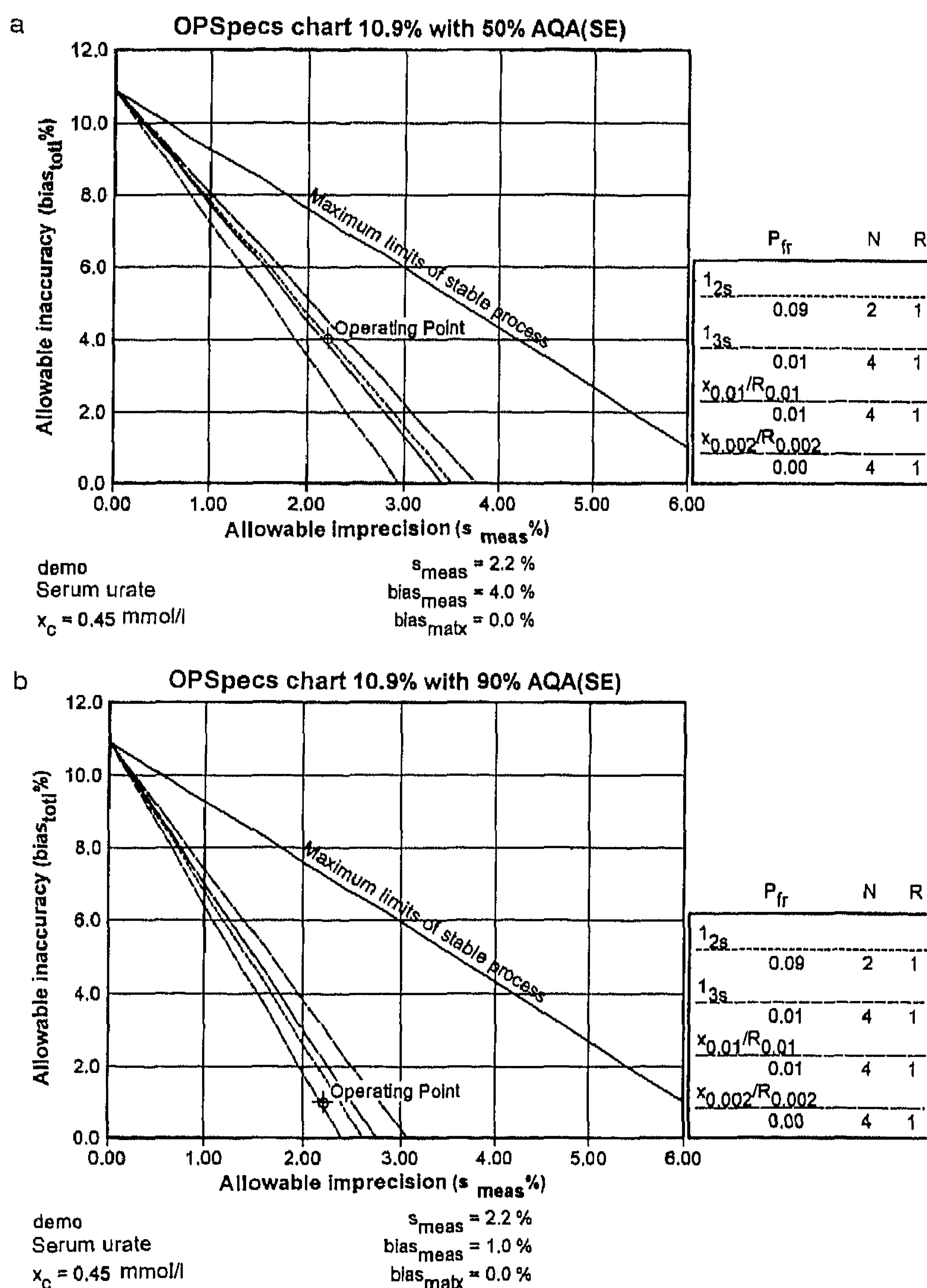


Fig. 11 OPSpecs Chart for serum urate with 'Allowable Total Error' equal to 10.9% (allowable bias = 4% and allowable imprecision = 4.2% {4% + 1.65 × 4.2%}), for two different performance characteristics:

a. Inherent imprecision to 2.2% and analytical bias to 4.0%, which can only give a power of 50%.

b. Inherent imprecision equal to 2.2% and analytical bias to 1.0%, whereby the 90% power can be obtained, using the same control rules.

Deviations in P_{fr} - and P_{ed} -values from theory are due to the computer simulation technique. Print out from the QC Validator™-programme (see footnote³) with permission.

demo = example

s_{meas} = s measured

$bias_{meas}$ = bias measured

$bias_{matr}$ = bias due to matrix

tively. TE_{allow} is then equal to $13.6\% + 1.65 \times 13.6\% = 36\%$. Using the same four control rules as used in figures 8 and 9, and assuming an inherent imprecision of 3.5%, then the OPSpecs Chart from the Validator programme gives a plot as shown in figure 10.

The axes are allowable bias and allowable imprecision, respectively. The curve *Maximum limits of stable process* illustrates the possible combinations of allowable bias and imprecision and the four other curves illustrate the four control rules, which will guarantee this with $P_{\text{ed}} = 90\%$ (shown above in fig. 10). The *Operating Point* illustrates a situation for a laboratory with a bias equal to the maximum allowable (13.6%) and an inherent imprecision of 3.5%. This point must be to the left of the lines for control rules in order to guarantee the quality with the probability indicated. For this analytical component this is obtained for all four control rules. On the right the control rules are described.

2. For serum urate, the European analytical quality specifications are more demanding (26) with an allowable bias of 4.0% and an allowable imprecision of 4.2%. For a laboratory with a bias equivalent to the allowable level and an inherent imprecision of 2.2%, the quality cannot be guaranteed with a $P_{\text{ed}} = 90\%$, but only with 50%, and only for two control rules, the $\bar{x} \pm 2s$ with two replicates of the control, and with the combination of the two control rules $x_{0.01}$ and $R_{0.01}$ for four replicates of the control, but very different in P_{fr} (fig. 11a).

If, however, the analytical bias is reduced to 1%, then the power can be obtained for both combinations of mean and range rules and for the 1_{2s} -rule (with a P_{fr} equal to 9%) as illustrated in figure 11b.

It should be noted that the combination of mean and range rules, each with very low P_{fr} of 0.2%, with combined P_{fr} of 0.4%, are nearly as powerful as the other control rules with higher P_{fr} . This means that it is possible to keep the probability of false rejects low and at the same time have a high probability of error detection. By use of such a control system, there will be no reason for performing re-runs as a false reject signal will be seen only once a year, and in consequence, any reject signal means that there is an error and trouble-shooting should be activated. However, such control rules are best applied to batch oriented analytical systems rather than the dynamic random access approaches favoured at present.

Control at Two or More Concentrations

Until now we have dealt with only one control material where we have assumed that the concentration was close to the 'critical concentration'. The number of concentrations to be controlled depends on the use of the quantity,

and when there are more 'critical concentrations', each should be treated in the same way as described above.

When more materials are used, then the P_{fr} of a run is changed as the probability of false rejections are independent. The formula combining these P_{fr} -values to a total $P_{\text{fr}}(\text{total})$ is

$$P_{\text{fr}}(\text{total}) = 1 - [1 - P_{\text{fr}}(1)] \times [1 - P_{\text{fr}}(2)] \dots$$

where the P_{fr} -values are given in fractions. For small values of P_{fr} , however, the error of simple addition is negligible, e. g. $P_{\text{fr}}(\text{total}) = 0.01 + 0.002 = 0.012$.

For P_{ed} the combined effect of two or more controls is more difficult to interpret, and combinations should be performed with caution. First, the P_{ed} -values for random and systematic errors cannot be combined (as was the case also for one control). Second, the errors at two (or more) concentrations may be independent or even reverse, as seen for RIA curves, where the effect of errors on the S-shaped curve may be positive at one concentration and negative at another. Therefore, P_{ed} -values can only be combined if they reflect the same error, which cannot be expected in general.

Often control materials at different concentrations have different purposes, and the results can be treated differently. In some assays the 'high control' is more of a check of sufficient concentration of a certain reagent, e. g. serum lactate dehydrogenase, where it may control the required concentration of NADH. In such a case replicates are seldom needed and the demands may be more loose than at the critical concentration. In consequence, a lower P_{ed} can be accepted and the P_{fr} can be kept low, e. g. $< 0.1\%$.

It is difficult to recommend how many control materials should be used, over which concentrations, and how often. An overall guideline could be to control the critical concentration or concentrations with control rules and replicated aiming for a high P_{ed} for critical errors. The power, however, should be related to the stability of the analytical system and expected frequency of errors. Moreover, other concentrations as the high for which samples should be diluted and low, about the detection limit, if this has clinical relevance, may be controlled less intensely, e. g. once a run or even less, dependent on the frequency of changes of reagents and other key elements.

Other Control Systems

For some types of analytical work and some analytical principles and instruments, the control systems and control rules should be modified in order to optimize the system according to the special problems and to avoid too stringent control systems.

1. Control of fully automated single channel systems

These types of instruments are generally very stable with good precision and the possible errors for these types of systems are mainly related to batch-to-batch variations as the procedures in the laboratory are often reduced to a minimum. The principle control should be concentrated on each new batch, which should be evaluated carefully before it is introduced. The usual control rules can be used for this purpose, but often the general statistical evaluations will be more useful as all results can be investigated together.

Many instruments have several 'fail-safe' procedures built in, which will stop the operation, when certain errors occur, such as 'unacceptable calibration curve', 'wrong reagents', 'reverse rack', etc. For such instruments, the control systems may be loosened regarding P_{ed} , but the P_{fr} must be kept low (if self-inflicted problems are to be avoided).

2. Control of random access and multichannel instruments

For these instruments special attention should be paid when starting up (e.g. in the morning) using, e.g., two to four control measurements before running the patient samples (36). Then Cusum rules could be used during the production, for each quantity or in combination and combined with other control rules. Here, it is crucial to keep P_{fr} extremely low in order to avoid false reject signals from the repeated testing (cf. tab. 2).

In some instruments, the process is identical for several different quantities, so control results from all these can be combined e.g. in a Cusum-plot after 'normalization' of results. Thus, the number of control samples for each analysis can be reduced — often to only one control sample per quantity per day — but the frequency of control should be adjusted according to the stability, i.e. the type and frequency of errors.

3. Near patient testing

For the type of instruments to be used outside the main laboratory, e.g. dry chemistry systems, the quality depends on the analyte, the users skills (sampling and handling of the blood samples), and the instrument design. The problems related to the chemical part should be handled separately and the control should be performed accordingly. Regarding the sampling and measuring process, however, the control is very difficult. The control of imprecision may be performed with lyophilized materials, whereas control of systematic errors/bias often need matrix correct materials (37). Therefore, education of the nurses or others involved in testing is essential. Real control can only be performed by simultaneous blood drawing by the nurse and a laboratory

technician, or by using a 'stable' person for the control, e.g. by sampling and measuring blood from the same healthy person once a month (38). The principle is that for many biological quantities, the within-subject biological variation is rather low for healthy individuals, even for blood glucose (fasting). By use of a 'stable' person (living control) it is possible to monitor the combined variation of within-subject biology, sampling, handling of the specimen, and the analytical process. For quantities with moderate within-subject biological variation mean-values from more than one 'living control' could be used. The process, however, is time consuming and is only useful for occasional control, e.g. once a month.

4. Control of unstable quantities

For unstable quantities, e.g. haematological analyses, a short-term control procedure can be obtained by fresh stabilized blood, which can be used as long as the components is stable. If the in vitro disappearance rate is known, as for fractions of reticulocytes, the control material may be used for longer periods, as the disappearance is predictable. Since the patient results are the final products and they make up the majority of analytical outcome, it might be tempting to extract information about analytical quality from these data (39–41). The validity of patient results in analytical control as a substitute for control samples may be disputed, but median and distribution of patient results may be used as a final check of the production of larger series. In principle the median of patients results can be used after principles comparable to the mean-rules describes above, but it may be more difficult to give a reliable estimate of the standard deviation to be used. When stable and reliable control materials are difficult to obtain, however, the use of patient results in control of the analytical quality may be superior. Bull et al. (42) have evaluated mean-rules to be used in haematology and Cembrowski & Westgard (43) have assessed these control systems in comparison with traditional IQC, with results in favour of the patient control system. It must be remembered, however, that the 'quality of the patient data' must be documented. But with the increasing power of computers in laboratory medicine the patient data for such a control may make it valid for internal control of many analytical procedures.

5. Control of components measured on an ordinal- or nominal-scale

The control is often more simple but also less convincing from a statistical point of view. In practice two control samples are often run, one negative and one positive. For genetic mutants (DNA analyses) this is sufficient, but for many other analytes, e.g. HIV-antibodies, the more reliable control might be about the detection

limit. If a control with concentration near to the detection limit should be used, then the percentage of positive results should be a known value dependent on the definition of the detection limit and acceptance of dispersion. A great number of measurements will be needed to estimate this percentage.

Retrospective Documentation of Analytical Quality

In most laboratories, control results are also used for calculation of the analytical imprecision for each method. This technique implies, however, a sampling bias, as the results have first been used in the control procedure (this can partly be avoided if all control data are considered, including those rejected). An independent estimate can only be performed by independent documentation, e. g. use of a material which is measured as a patient sample, but not used in the control process. All results for this material from accepted runs can then be used for calculation of the analytical imprecision.

Relation to External Quality Assessment

As stated above, internal control is related to the variable factors, cannot in general be used for assessing 'traceability' (19, 20) and can only be used for control of specificity by use of combination of samples, selected for the purpose. It is, however, important to tie the internal control to the external system in order to make it reliable for monitoring of the long-term stability with a known and accepted bias. Therefore, target values should be assigned to the internal control materials when the traceability is assessed, by use of (international) reference preparations, external assessment using high quality materials with traceable target values (only known retrospectively), e. g. by running both types of materials in several runs using replicates. It is stressed that the control materials must be independent of the producer of kits and calibrators in use as independency is essential for reliable internal control. A scheme for integration of internal control and external assessment should be designed with number of replicates for internal as well as external materials related to the analytical quality specifications and confidence intervals estimated from the known inherent analytical imprecision. This increases the demands for reliable external control materials with target values traceable to internationally accepted certified reference materials or reference methods.

Relations to Other Recommendations

Guidelines for internal control have been published for the concentration levels of control materials, the number of replicates of each material, and control rules to be

used (44). NCCLS (45) has defined that 'quality control samples must be analyzed at least once during each user-defined analytical run' and the German "Richtlinien" (46, 47) require precision control at 'normal range' in every 4th series. The purpose of 'Richtlinien' and thereby, the German control system is a system with traceable values which in co-ordination with a firm internal control system is believed to secure a high level of analytical quality. IUPAC (48) recommends control charts of the *Shewhart* type and also deals with analysis in duplicate.

In principle we agree with all the recommendations from these guidelines. We find, however, that they point too little to the purpose of improving the analytical performance and prevention of errors, which must be more interesting than the control. Here, the IQC is an indispensable part — but only a part — of the whole process as described in figure 1. In consequence, we do not recommend a certain number of control materials and a defined number of replicates. In contrast, the control system should be flexible and related to the current stability (or lack of stability) of the analytical system.

As mentioned above control does not improve analytical quality by itself, but is useful in detection of errors. When errors occur, however, it is important to do the trouble-shooting, and — not only correct the error — but investigate the cause, and finally make systems to prevent the same error from occurring again (fail-safe procedures). Therefore, a control system should not be static, but dynamic, and one prerequisite is a control system with a very low P_{fr} , in order to reduce frustrations from unnecessary reject signals.

Guidelines

The internal control system should be problem related and used as an integrated part of a more dynamic system for quality improvement.

It should be linked to the external system with reliable control materials with concentrations 'traceable' to internationally accepted certified reference materials or reference methods, through a design which guarantees the further 'traceability' in the laboratory.

It should further be related to the analytical quality specifications, e. g.

$$\begin{aligned} CV_{\text{analytical}} &< 0.5 \times CV_{\text{within-subject}} \text{ and} \\ \text{Bias}_{\text{analytical}} &< 0.25 \times \\ &\quad (CV_{\text{within-subject}}^2 + CV_{\text{between-subject}}^2)^{1/2} \end{aligned}$$

When errors are frequent the P_{ed} should be high, but with a very low P_{fr} , in order to stimulate intensive trouble-shooting and build up a fail-safe system, which will lower error frequency, and thereby, open for a looser control system, but still with a low P_{fr} .

Other control systems and check procedures (built into the equipment or information from other analytes measured on the same equipment) should be integrated into the total control system. Where possible, median and distribution of patient results should be taken into consideration (e.g. haematology where the indices are rather constant), but it must be remembered that patient results can give supplementary information, but they cannot replace all control samples.

Control charts of the *Shewhart* type should be shown on the control screens or printed out for visual inspection for minor systematic errors and supplemented with a Cusum-rule or comparable.

It is better to prevent errors than to correct them and to reduce imprecision, which will make it easier to combine a low P_{fr} with a high P_{ed} .

References

1. Levey S, Jennings ER. The use of control charts in the clinical laboratory. *Am J Clin Pathol* 1950; 20:1059–66.
2. Shewhart WA. Economic control of quality of the manufactured product. New York NY: Van Nostrand, 1931.
3. Westgard JO, Groth T, Aronsson T, Falk H, de Verdier C-H. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. *Clin Chem* 1977; 23:1857–67.
4. Westgard JO, Groth T, Aronsson T, de Verdier C-H. Combined Shewhart-Cusum control chart for improved quality control in clinical chemistry. *Clin Chem* 1977; 23:1881–7.
5. Westgard JO, Groth T. Power functions for statistical control rules. *Clin Chem* 1979; 25:863–9.
6. de Verdier C-H, Aronsson T, Nybnerg A, editors. Quality control in clinical chemistry – efforts to find an efficient strategy. Helsinki, Finland: The Nordic Clinical Chemistry Project (NORDKEM), 1984.
7. Westgard JO, Barry PL. Cost-effective quality control: managing the quality and productivity of analytical processes. Washington, DC 20006: AACC Press, 1725 K Street, NW, 1986.
8. Cembrowski GS, Carey RN. Laboratory quality management QC \leftrightarrow QA. Chicago: ASCP Press, American Society of Clinical Pathologists, 1989.
9. Linnet K. Choosing quality control systems to detect maximum clinically allowable analytical errors. *Clin Chem* 1989; 35:284–8.
10. Linnet K. Mean and variance rules are more powerful or selective than quality control rules based on individual values. *Eur J Clin Chem Clin Biochem* 1991; 29:417–24.
11. Hyltoft Petersen P, Blaabjerg O, Irjala K, Icen A, Bjoro K. Elements of analytical quality. In: Hyltoft Petersen P, Blaabjerg O, Irjala K, editors. Assessing quality in measurements of plasma proteins. The Nordic Protein Project and Related Projects. NORDKEM. Helsinki, Finland: Nordic Clinical Chemistry Project, 1994:19–39.
12. Stöckl D, Baadenhuijsen H, Fraser CG, Libeer J-C, Hyltoft Petersen P, Ricós C. Desirable routine analytical goals for quantities assayed in serum. Discussion paper from the members of the external quality assessment (EQA) working group A on analytical goals in laboratory medicine. *Eur J Clin Chem Clin Biochem* 1995; 33:157–69.
13. Hyltoft Petersen P, Lytken Larsen M, Hørder M, Blaabjerg O. Influence of analytical quality and preanalytical variations on measurements of cholesterol in screening programmes. *Scand J Clin Lab Invest* 1990; 50 Suppl 198:66–72.
14. Hyltoft Petersen P, de Verdier C-H, Groth T, Fraser CG, Blaabjerg O, Hørder M. On the influence of analytical bias on diagnostic misclassifications. *Clin Chim Acta*, 1996. In press.
15. Libeer J-C. External quality assessment in clinical laboratories. Thesis, Antwerpen, 1993.
16. Hyltoft Petersen P, Blaabjerg O, Irjala K, Icen A, Bjoro K. External analytical quality assurance for proteins. In: de Verdier C-H, Groth T, Hyltoft Petersen P, editors. Medical need for quality specifications in clinical laboratories. *Upsala J Med Sci* 1993; 98:241–58.
17. Libeer J-C, Baadenhuijsen H, Fraser CG, Hyltoft Petersen P, Ricós C, Stöckl D, Thienpont L. Characterization and classification of external quality assessment schemes (EQA) according to the objectives such as examination of method and participant bias and standard deviation. *Eur J Clin Chem Clin Biochem* 1996; 34:665–78.
18. Thienpont L, Franzini C, Kratchovila J, Middle J, Ricós C, Siekman L, Stöckl D. Analytical quality specifications for reference methods and operating specifications for networks of reference laboratories. Discussion paper from the external quality assessment (EQA) working group B on target values in EQAS. *Eur J Clin Chem Clin Biochem* 1995; 33:949–57.
19. BIPM, IEC, IFCC, ISO, IUPAC, OIML. International vocabulary of basic and general terms in metrology, 2nd ed. Geneva: ISO, 1993:1–59.
20. Dybkær R. Vocabulary for describing the metrological quality of a measurement procedure. *Upsala J Med Sci* 1993; 98:445–86.
21. Hyltoft Petersen P, Blaabjerg O, Irjala K, Icen A, Bjoro K. The Nordic Protein Project. In: Hyltoft Petersen P, Blaabjerg O, Irjala K, editors. Assessing quality in measurements of plasma proteins. The Nordic Protein Project and Related Projects. NORDKEM. Helsinki, Finland: Nordic Clinical Chemistry Projekt, 1994:87–116.
22. Westgard JO, Hyltoft Petersen P, Groth T. The quality-cost of an analytical process. I. Development of quality-costs models based on predictive value theory. *Scand J Clin Lab Invest* 1984; 44 Suppl 172:221–7.
23. Westgard JO, Hyltoft Petersen P, Groth T. The quality-cost of an analytical process: II. A test yield formulation of the predictive value quality-costs model. *Scand J Clin Lab Invest* 1984; 44 Suppl 172:229–36.
24. Ricós C, Baadenhuijsen H, Libeer J-C, Hyltoft Petersen P, Stöckl D, Thienpont L, Fraser CG. External quality assessment: currently used criteria for evaluating performance in EQA in European countries and criteria for further harmonization. (Discussion paper from the members of the EQA working group A on analytical goals in laboratory medicine). *Eur J Clin Chem Clin Biochem* 1996; 34:159–65.
25. Blaabjerg O, Blom M, Gry H, Hyltoft Petersen P, Uldall A. Appropriate sera for calibration and control of specific protein assays. *Scand J Clin Lab Invest* 1993; 53 Suppl 212:13–5.
26. Middle J, Blaabjerg O, Libeer J-C, Penttilä I, Malakhov NV. European EQAS organizers working group C. Characterisation of evaluation of quality control materials. In preparation.
27. 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* 1980; 26:394–400.
28. Fraser CG, Hyltoft Petersen P, Ricós 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.
29. Westgard JO, Quam ES, Barry PL. Selection grids for planning quality control procedures. *Clin Lab Sci* 1990; 3:271–8.
30. 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.
31. 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.
32. Westgard JO, Seehafer JJ, Barry PL. Allowable imprecision for laboratory tests based on clinical and analytical test outcome criteria. *Clin Chem* 1994; 40:1909-14.
 33. Westgard JO, Seehafer JJ, Barry PL. European specifications for imprecision and inaccuracy compared with operating specifications that assure the quality required by US CLIA proficiency-testing criteria. *Clin Chem* 1994; 40:1228-32.
 34. Skendzel LP, Barnett RN, Platt R. Medically useful criteria for analytical performance of laboratory tests. *Am J Clin Pathol* 1985; 83:200-5.
 35. US Dept. of Health and Human Services. Medicare, Medicaid, and CLIA Programs; regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Final rule. *Fed Regist* 1992; 57:7002-186.
 36. Aronsson T, Groth T. Nested control procedures for internal analytical quality control. Theoretical design and practical evaluations. In: de Verdier C-H, Aronsson T, Nyberg A, editors. *Quality control in clinical chemistry - efforts to find an efficient strategy*. Helsinki, Finland: The Nordic Clinical Chemistry Project (NORDKEM), 1984:51-64.
 37. Juvany R, Ricós C. Can solid and liquid phase chemistry analyzers be submitted to quality control by the same lyophilized material? *Quim Clin* 1994; 13:243-9.
 38. Hyltoft Petersen P, Birkemose E, Pedersen B, Jørgensen PJ, Grinsted P, Wang P, Hørder M. The living control: improved QC of hemoglobin analysis in the practitioners clinic. *Clin Biochemist-Reviews* 1993; 14:344.
 39. Hofman RG, Waid ME. The 'average of normals' method of quality control. *Am J Clin Pathol* 1965; 43:134-41.
 40. Cembrowski GS, Chandler EP, Westgard JO. Assessment of "average of normals" quality control procedures and guidelines for implementation. *Am J Clin Pathol* 1984; 81:492-9.
 41. Cembrowski GS. Use of patient data for quality control. *Clin Lab Med* 1986; 6:715-33.
 42. Bull BS, Elashoff RM, Heilbron DC, Couperus J. A study of the various estimators for the derivation of quality control procedures from patient erythrocyte indices. *Am J Clin Pathol* 1974; 61:473-81.
 43. Cembrowski GS, Westgard JO. Quality control of multichannel hematology analyzers: Evaluation of Bull's Algorithm. *Am J Clin Pathol* 1985; 83:337-45.
 44. Büttner J, Borth R, Boutwell JH, Broughton PMG, Bowyer RC. International Federation of Clinical Chemistry. Approved recommendation on quality control in clinical chemistry. Part 4. Internal quality control. *J Clin Chem Clin Biochem* 1983; 21:877-84.
 45. Approved guideline. Internal quality control testing: principles and definitions. NCCLS Document C24-A, Vol. 11, No. 6, 1991.
 46. Richtlinien der Bundesärztekammer zur Qualitätssicherung in medizinischen Laboratorien. *Dtsch Ärztebl* 1988; 85:A699-712.
 47. Reinauer H. External quality assessment schemes for clinical chemistry in Germany. *Ann Ist Super Sanità* 1995; 31:77-86.
 48. Thompson M, Wood R. Harmonised guidelines for internal quality control in analytical chemistry laboratories. International Union of Pure and Applied Chemistry, 1993.

Received May 27/September 26, 1996

Corresponding author: Dr. Per Hyltoft Petersen, Klinisk Biokemisk Afdeling, Odense Universitetshospital, DK-5000 Odense C, Denmark