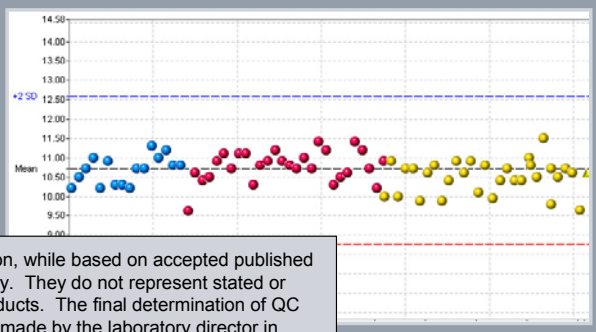
SIEMENS

Designing Effective QC

Making QC effective and efficient by design

Nils B. Person, Ph.D., FACB
Senior Scientist
Global Product Education



NOTE: Recommendations made in this presentation, while based on accepted published guidelines and literature, are recommendations only. They do not represent stated or implied requirements for operation of Siemens products. The final determination of QC protocols and procedures used in the laboratory is made by the laboratory director in compliance with all applicable regulatory requirements.

Welcome to the Siemens Healthcare Educational session on Designing Effective QC. For the next 90 minutes we are going to look at some of the tools available to design a laboratory QC protocol that can reliably detect significant change in method performance while also being cost effective and practical. Probably more cost effective and practical that what is commonly done in many laboratories today.

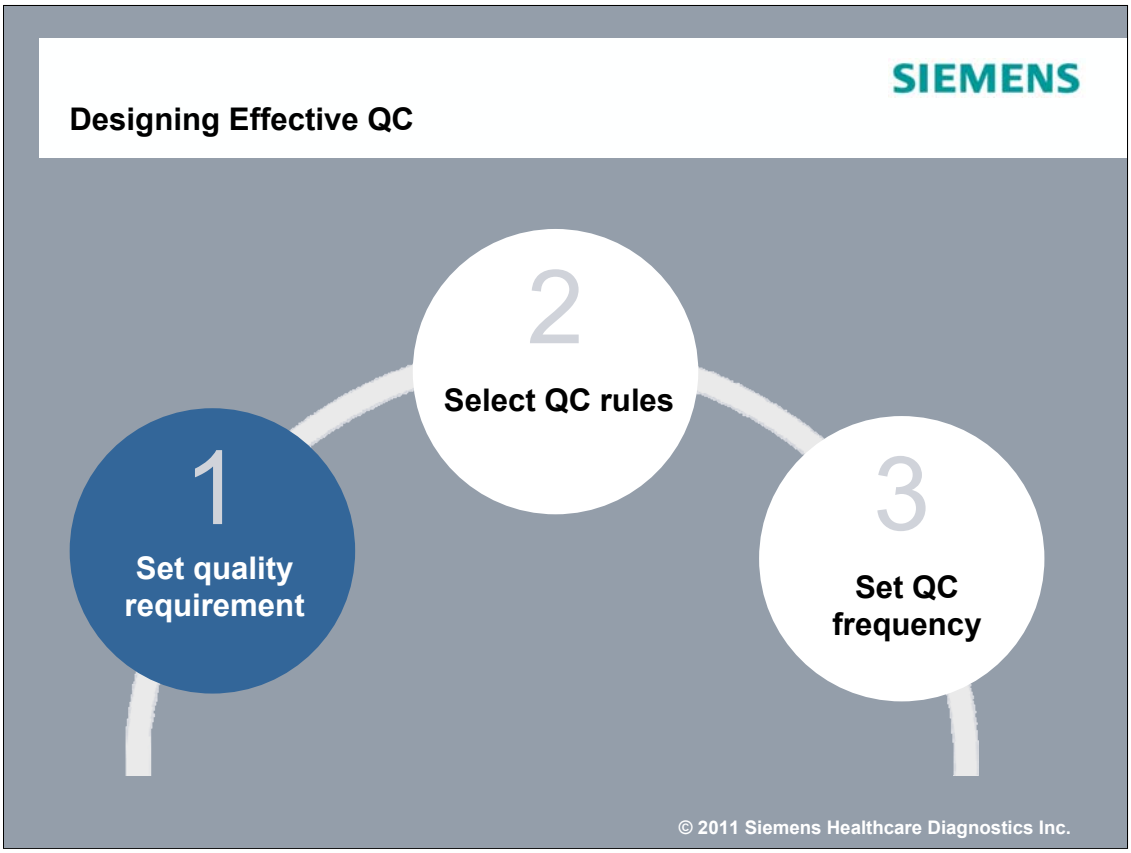
SIEMENS

Objectives

- List steps to designing effective QC protocol
- State how Total Allowable Error can be used to optimize QC procedures
- List key points in using QC rules effectively

Page 2

Our objectives for this session are to look at the steps involved in designing effective QC, to look at how Total Allowable Error can be effectively used in the process and to review some of the key points to keep in mind when using QC rules to enhance effectiveness.



There are three steps to designing effective QC. First set the quality requirement. Next Select the QC rules that will be most efficient at meeting that requirement, and then finally determining how often we need to test QC samples to be efficient and effective.

Let's start by discussing the quality requirement.

SIEMENS

When is there a performance problem?

QC results fail a statistical rule

Tool

≠

Method performance has changed enough to impact medical care

Quality Requirement

7. The supervisor may make a decision to report data when there is a lack of statistical control in the following situations:
   (c) The control problem occurs in a concentration range that is different from the concentrations of the patients' samples. The method is in-control in the range of the patients' samples.
   (d) The size of the analytical error is judged to be small relative to the medical usefulness requirements.[3]

Note: Reviewer R.B. noted that this list (in 7 above) should not

[3] The quality of laboratory service is related not only to the size of analytical errors, but also to other factors such as the time required to obtain the result. There may be situations that require a relative judgment of the importance of the various factors involved in quality, thus it may be necessary to overrule the statistical control system. This is a professional judgment requiring knowledge of medical usefulness limits of error, an understanding of the use and interpretation of the analytical results, and experience.

Is that always true ?

Westgard et al, A Multi-Rule Shewart Chart for Quality Control in Clinical Chemistry, Clin Chem, 27, 493, 1981

Page 4

A key concept to think about when looking at QC results is what constitutes a real performance problem ? How do we know when there is a real meaningful problem with the method ?

Is it always meaningful when the QC results fail a statistical QC rule ? Or is the true criterion that method performance has changed enough to impact medical care. I think we all agree that the later is our real concern. However, historically we have mostly all assumed that these two things were equal and identical; that failure of a statistical QC rule **always** meant that there was a medically significant change in method performance. But is that always the case ? Actually, we know it is not. This is reflected in the fact that we often report results even though the QC results have failed a rule. We say that if only one level is out, it's OK to report, or similar evaluations. Here we have an excerpt from the original "Westgard Rules" paper in 1981 that indicates that it is acceptable and necessary to sometimes recognize that just because QC results have failed a statistical rule, even the "Westgard Rules", it may still be reasonable and necessary to release results while we are investigating the rule failure.

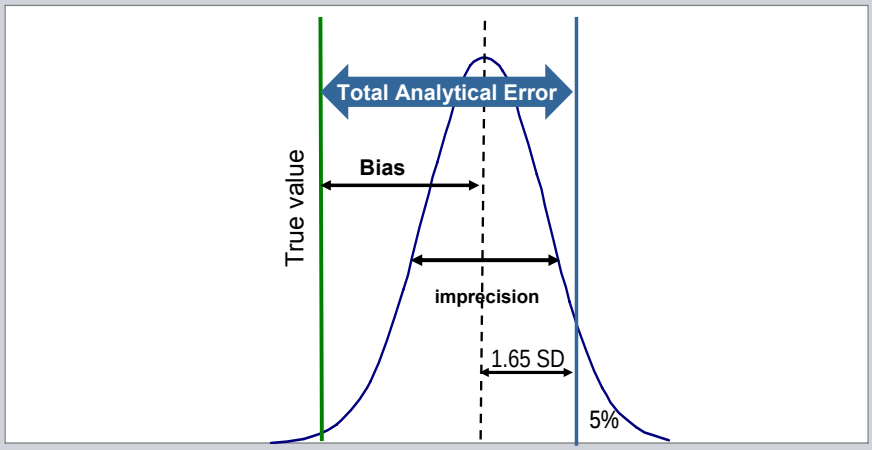
Statistical QC rules are tools we use. Tools that help us to know that some degree of change has occurred in the method. Then that change needs to be put in perspective relative to our quality requirement for method performance. We have been doing this intuitively and informally for years.

Let's look at formally establishing our quality requirement and see how QC rules really relate to it. To do this we want to introduce a two very useful concepts – Total Analytical Error and Total Allowable Error.

SIEMENS

Total Analytical Error

Error that encompasses 95% of results



Total Analytical Error = bias + 1.65(SD)

Total analytical error is defined as the error that encompasses 95% of the results for a given method. As we know, error is made up of two components. Constant error, often called bias, is the average consistent error seen over time. We would like to reduce or eliminate bias, but cannot always do so. The other component of total error is random error or imprecision. This is an inherent characteristic of the method. To estimate the total error for the method we want to capture the error that covers 95% of the results of the method. Since random variation is symmetrical around the mean, it sometimes adds to total error and sometimes reduces total error. Since we are only interested is the maximum error, we only look at the random error that increases total error. Using the SD as the measure of random error, the combined bias and imprecision that covers the error for 95% of results is bias plus 1.65 times the SD. That becomes our formula for estimating total analytical error.

SIEMENS

Total Analytical Error

-the actual error that the method normally has

Total Allowable Error

- the maximum error before something bad happens

Together can be used to select optimal QC protocol

Page 6

So total analytical error is the actual error we have. We next want to look at what is the maximum error that can be tolerated before we impact patient care. There have been a number of ways to describe this, but the one that is currently most widely used is Total Allowable Error. Using these two concepts together ca help us design effective and efficient QC. Let's look at Total Allowable Error in more detail.

SIEMENS

Total Allowable Error

Total Allowable Error (TE_a): Maximum error that can be tolerated before some outcome is affected

Not method performance based

Based on how results are used, not generated

Determined by change in outcome

Outcomes:

- Failed PT
- Altered medical decision
- Altered patient care

Established at decision points

Examples:

Glucose	- 120 mg/dl
Na+	- 115 mmol/L
PSA	- 4.0 ng/dl
TSH	- 4.0 µIU/ml

Total Allowable Error is the maximum error we can tolerate for an assay before some outcome like medical decision making or patient care is impacted.

Total allowable error is NOT based on current method performance. It is determined by how the results are used medically, not how the results are determined analytically. So it is independent of the method used. Since Total Allowable Error is dependent on the clinical use of the test result and the inherent biologic variability of the analyte, it is not the same for all analytes. Therefore it has to be established for each analyte and for each medically important concentration for the analyte. The total allowable error for calcium is the same regardless of what instrument or method is used to measure calcium.

The idea of total allowable error is that if we exceed it, the some outcome will be affected ... we may fail proficiency testing, a medical decision may be altered, patient care may be changed.

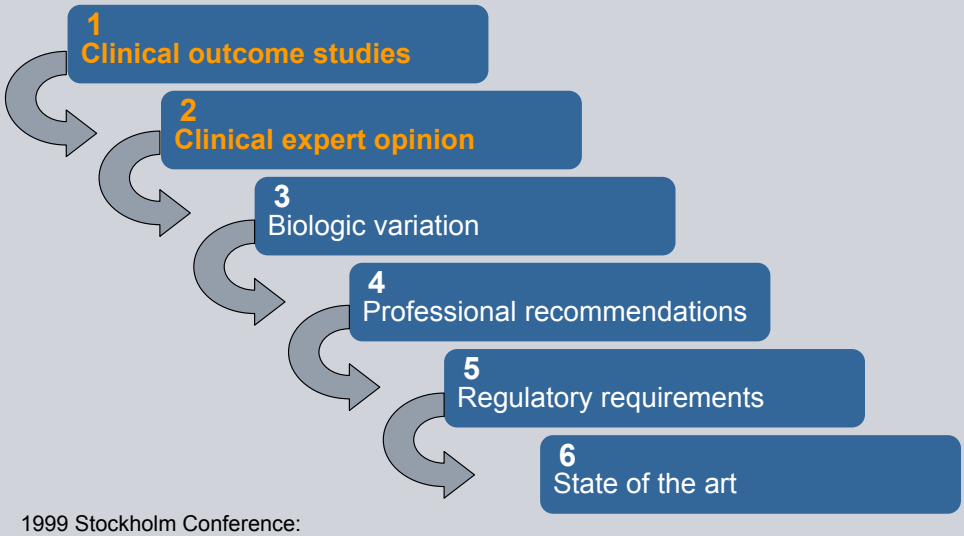
Since the concept of total allowable error revolves around medical decision making, typically we estimate the allowable error at concentrations where medical decision are made. To understand how this concept may be used let's try defining Total Allowable Error for a few specific analytes as examples. We'll use Glucose, Sodium, PSA and TSH. The first step for each analyte is to define a medically important concentration. For Glucose, 120 mg/dl is a decision point for the diagnosis of diabetes; for Sodium 115 mmol/L is the decision point for hyponatremia and severe electrolyte imbalance; for PSA a result above 4 ng/dl is suggestive of increased risk for cancer and should be followed up; and for TSH a result above 4.0 uIU/ml indicates possible hypothyroidism.

So, how do we decide what Total Allowable Error should be for a method ? Many authorities discussed this for a number of years and in 1999 there was a conference held is Stockholm to develop a consensus approach.

SIEMENS

Determining Total Allowable Error

Recommended Hierarchy for Specifications:



1999 Stockholm Conference:

Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Strategies to set global analytical quality specifications in laboratory medicine. Consensus agreement. *Scand J Clin Lab Invest.* 1999;59:585

Page 8

At the conference it was recognized that there is not one simple approach that will work to define Total Allowable Error for all methods. So a hierarchy was developed to start with the most medically sound approaches and move to other approaches if the optimal was not possible. Here is that hierarchy.

Let's start with looking at the use of outcome studies and Clinical Expert opinion

Total Allowable Error:
Clinical Outcome & Expert Opinion

SIEMENS

How much change in a result alters medical outcome ?

- Becomes the total allowable error for that analyte

Clinical outcome studies:

- Cardiac disease – Framingham, TIMI, Women's Health Study
- Diabetes – DCCT, NHANES,
- Large, prospective, long term studies looking at clinical outcome

Expert Opinion:

- Review institutional standardized care protocols
- Consult with physicians for expert opinion

Page 9

We are trying to establish how much change in a result will alter medical decision making and patient care. That amount of change then becomes our allowable error since any change less than that will not cause a physician to make a different decision.

Clinical outcome studies are the optimal source for this information. They are focused on the decision making in specific medical scenarios, like diagnosis and management of heart disease or diabetes. These studies are prospective, long term studies that objectively assess how treatment decisions affect the outcome. Often lab results are used to make the treatment decisions. This is the most specific data we can use.

Using medical expert opinion may seem an obvious choice for setting Total Allowable Error. Essentially we want to know how much change in the result for a test will cause physicians to change their decision and that becomes the limit. To assess this we can look at the consensus derived standard treatment protocols that are used in many healthcare facilities today or we can consult with physicians. This will give us the benefit of their collective experience on how to best use lab results.

Total Allowable Error:
Clinical Outcome & Expert Opinion - Example

SIEMENS

HbA1c

Clinical outcome study:

DCCT: increase of HbA1c of 1% (i.e. HbA1c result going from 7% to 8%) leads to significantly poorer outcome

Expert opinion:

Endocrinologists indicate that they view a 10% change (i.e. HbA1c result going from 8% to 7.2%) indicating significant change in patient

Page 10

An example of a method with clinical outcomes-based data that can be used to make comparability recommendations is the use of the hemoglobin A1c assay (HbA1c) for monitoring an individual's diabetes control. The Diabetes Control and Complications Trial on Clinical Outcomes Related to HbA1c indicated that a HbA1c of 8.0% has a poorer clinical outcome compared to a HbA1c of 7.0%, and should therefore be accompanied by a change in patient management.

We can use the same example for Expert Clinical Opinion. A survey of endocrinologists might indicate that clinicians interpreted a 10% change (eg, a change in HbA1c concentration from 8.0% to 7.2%) in the HbA1c result as a significant change in a patient's clinical condition.[i]

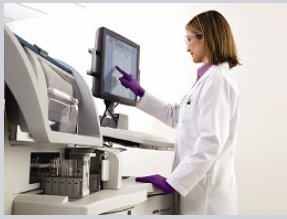
[i] Petersen PH, Larsen ML, Horder M. Prerequisites for the maintenance of a certain state of health by biochemical monitoring. In: Harris EK, Yasada T, eds. *Maintaining a Healthy State Within the Individual*. Amsterdam: Elsevier; 1987:147-158.

Total Allowable Error:
Clinical Outcome & Expert Opinion

SIEMENS

Challenges:

- Outcome study data do not exist for most analytes
- Standardized protocols often assume all lab results equivalent; do not state performance criteria
- Consistent results across methods / laboratories becomes critical
- Physician's intuitive sense of significant change influenced by historical variability of lab results
- May be conditioned by older laboratory technology



Page 11

The challenge is that outcomes studies and clinical protocols don't exist for most analytes. So, while they may be useful guidance for some analytes, for most there is no standard of this type. Also, when soliciting expert opinion, how do you decide how much change is critical ?

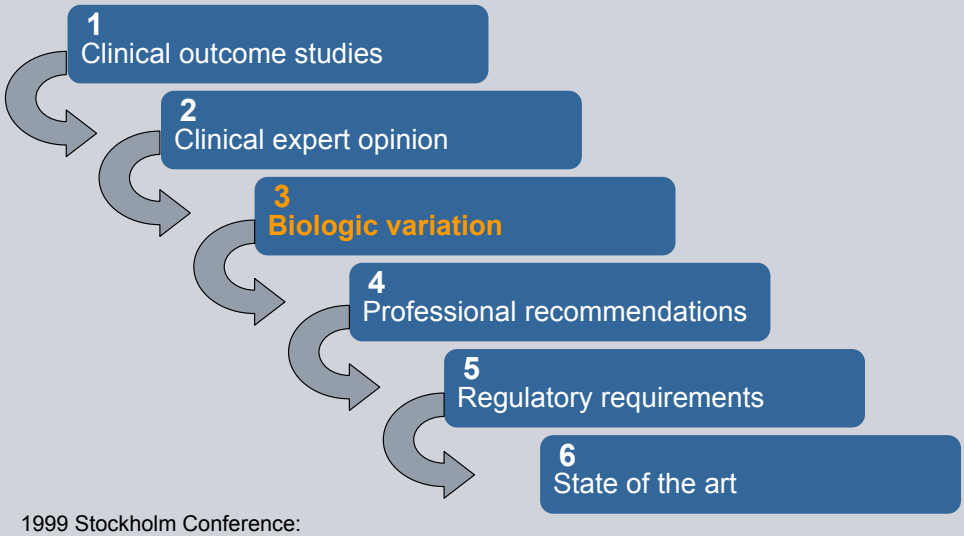
It may seem straightforward to just consult with physicians about how much a Glucose result has to change before they would consider it significant, but there's a problem. Physician's intuitive sense of how much change is significant is in large part based on their experience with how variable lab results are compared to changes noted in the patient's status. This intuitive sense is shaped by the variability in lab results seen in the past and doesn't necessarily reflect current test performance.

So these approaches are very valuable, but may not be practical for all analytes.

SIEMENS

Determining Total Allowable Error

Recommended Hierarchy for Specifications:

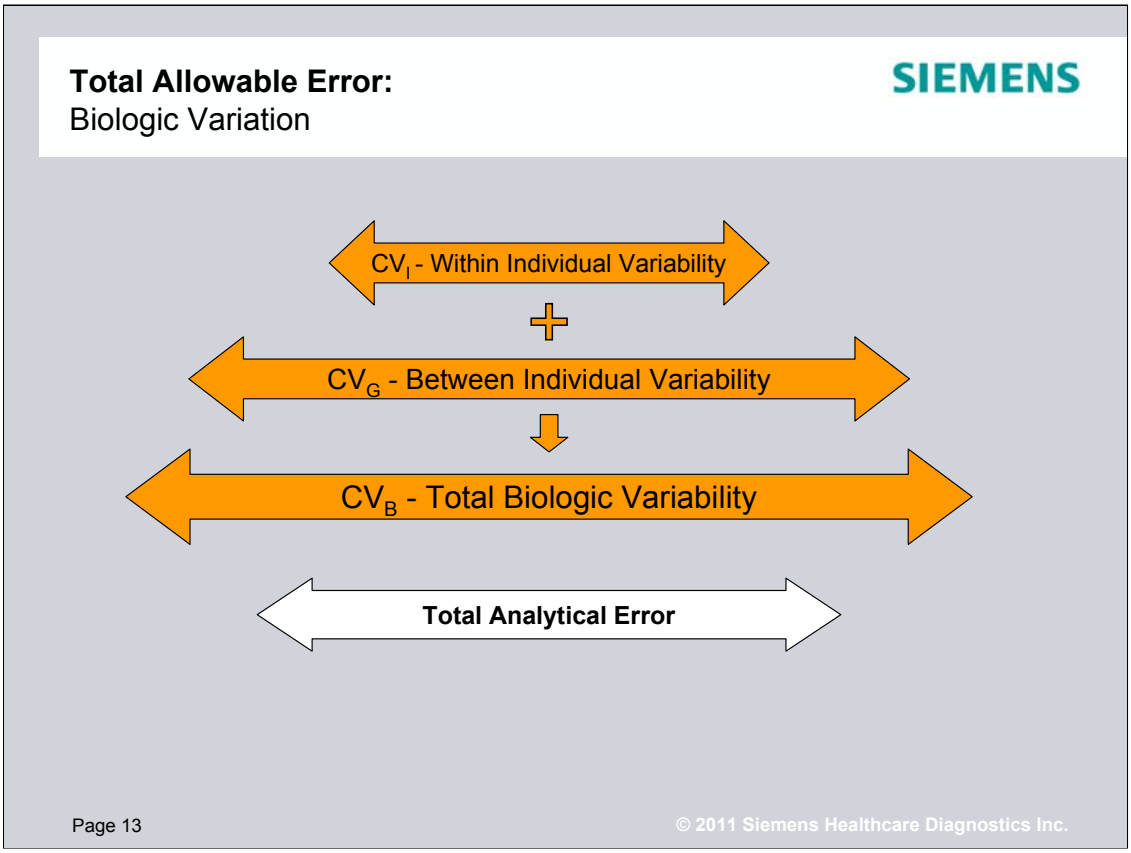


1999 Stockholm Conference:

Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Strategies to set global analytical quality specifications in laboratory medicine. Consensus agreement. *Scand J Clin Lab Invest.* 1999;59:585

Page 12

Now let's look at Biologic variation



All analytes we measure show inherent intra-individual variation. There is typically some degree daily variation that may follow a circadian rhythm. There are longer term variations including some seasonal. All of these variations are independent of any pathologic change and are part of normal physiology.

Further, the usual analyte concentration varies between individuals as well. As might be expected this is a larger variation than is seen within a single individual. The combination of intra-individual and between individual variation is called total biologic variation. We have all seen this variation reflected in the reference interval or "normal" range commonly used to interpret lab results. As most often used, the reference interval represents the central 95% of the range of values found in a population of healthy individuals.

The goal in using biologic variation to set total allowable error is that the analytical error should be small compared to the natural biologic variation. That way the analytical error will essentially be "lost" in the background noise. A number of articles have been published on how to achieve this and a consensus has emerged.

Total Allowable Error:
Biologic Variation

SIEMENS

Current consensus goals:

Desirable maximum bias = 25% of total biologic variability = **$0.25(CV_B)$**

Desirable maximum imprecision = 50% of within individual variability = **$0.5(CV_I)$**

Total Allowable Error goal:

$$TE_a = 0.25(CV_B) + 1.65(0.5CV_I)$$

Total Analytical Error = bias + 1.65(CV)

Page 14

The current consensus on using biologic data to set analytical performance goals sets the limits of analytical error based on the biologic data. The desirable goal for bias is no more than 25% of total biologic variability. The desirable goal for imprecision is no more than 50% of within individual variability. Using these proposed limits, we can set a goal for Total Allowable Error that will encompass 95% of results for a given analyte.. The estimated Total Allowable Error is the bias plus 1.65 times imprecision. This is the current working model for estimating total allowable error from biologic data.

Total Allowable Error:
Biologic Variation - Example

SIEMENS

Desirable Specifications for Total Error, Imprecision, and Bias, Derived from Biologic Variation

Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, Minchinela J, Perich C, Simon M. "Current databases on biologic variation: pros, cons and progress." Scand J Clin Lab Invest 1999;59:491-500.

Annex I, Part I: Within-subject and between-subject CV values of analytes and *Desirable Analytical Quality Specifications for imprecision, bias and total error*

Analyte	Biologic Variation		Desirable Specification		
	CV_I (%)	CV_B (%)	CV (%)	Bias (%)	TE_a (%)
Glucose	5.7	6.9	2.9	2.2	6.9
Na^+	0.7	1.0	0.4	0.3	0.9
PSA	18.1	72.4	9.1	18.7	33.6
TSH	19.3	19.7	9.7	6.9	22.8

Page 15

To look at some of these biologically based goals, an excellent resource is an ongoing series of articles published by Carmen Ricos and colleagues. The table of biologic data can be readily accessed at Dr. Westgard's website.

Here we can see the data for our four example assays with the Total Allowable Error goal listed in the right most column.

Total Allowable Error:
Biologic Variation

SIEMENS

Challenges:

- No complete agreement on biologically based goals
- Variability data for some analytes not robust
- Performance of some current methods cannot meet biologic goals

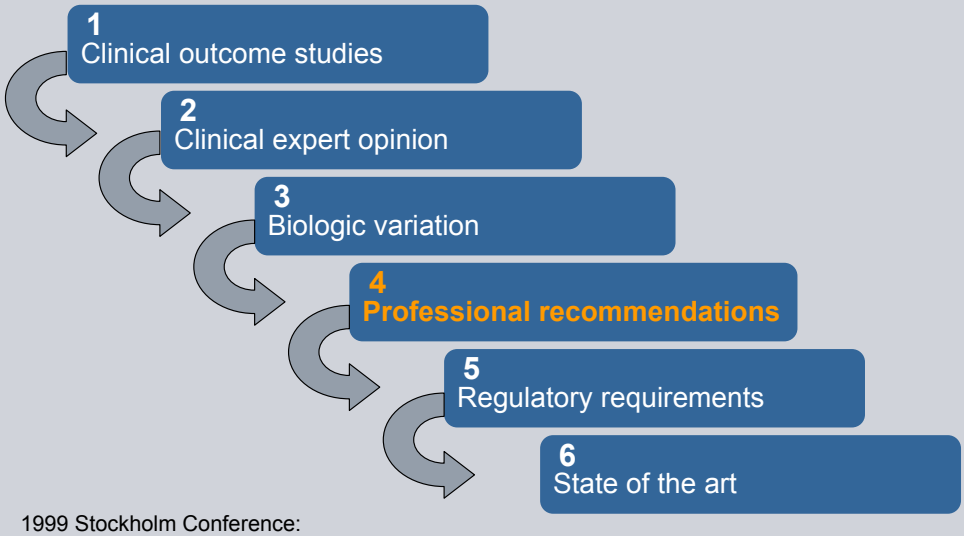
Page 16

There are challenges. First this is a consensus model, which implies some degree of disagreement on how the goals should be set. Second, the data used to determine biologic variability is not robust for all analytes. We have excellent data for many analytes, but the data is not as solid for many others. Finally, some methods in current use cannot achieve the level of performance necessary to meet goals set using this model. Current technology is not capable. Example analytes where this is an issue are Sodium and often Calcium.

SIEMENS

Determining Total Allowable Error

Recommended Hierarchy for Specifications:



1999 Stockholm Conference:

Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Strategies to set global analytical quality specifications in laboratory medicine. Consensus agreement. *Scand J Clin Lab Invest.* 1999;59:585

Page 17

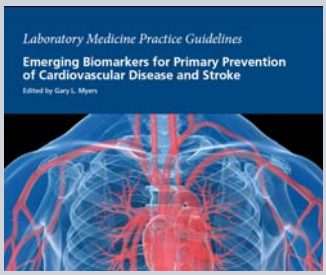
Another option is to use recommendation made by professional groups

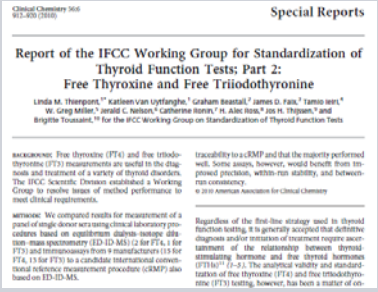
Total Allowable Error:
Professional recommendations

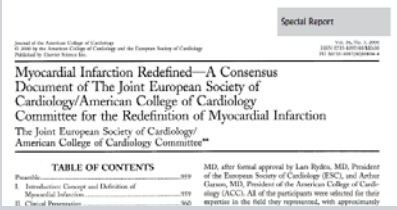
SIEMENS

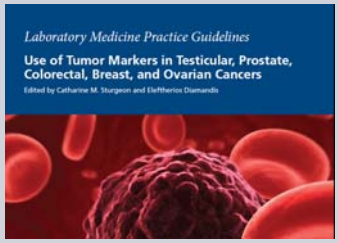
Professional group recommendations

Medical decision cutoffs & associated performance requirements









Page 18

There have also been a number of published studies and reports by professional groups that also establish specific medical decision points for some analytes. In these studies and reports, tolerable error limits are often also defined. These reports can be very useful in establishing Total Allowable Error for those analytes. Since the data used to establish the recommended performance criteria are not always outcome based, the recommendations in these reports are not a solid as those from outcome studies. These reports are based on outcome data whenever possible, but, as we already indicated, that data does not exist for many analytes.

Total Allowable Error:
Professional recommendations - Examples

SIEMENS

TE_a Goals based on Professional Recommendations:

- Cholesterol – NCEP - +/- 20% @ 200 mg/dl
- Glucose – ADA - +/- 25% @ 100 mg/dl
- TSH – NACB - +/- 40% @ 0.02 mIU/L

Page 19

Examples include the National Cholesterol Education Program (NCEP) establishing that cholesterol results changing by more than 20% at 200 mg/dl is clinically important, or the American Academy of Cardiology (ACC) indicating that the decision point for Troponin should be the 99th percentile of the healthy population and that the maximum allowable error at that concentration is 20%, or the National Academy of Clinical Biochemistry (NACB) publishing guidelines for thyroid testing that indicate that a change in TSH of more than 40% at 0.02 mIU/L is significant.

Total Allowable Error:
Professional recommendations

SIEMENS

Challenges:

- Published guidelines only cover limited number of analytes
- Standardized guidelines require consistency across methods / labs
- Some current methods cannot meet desired performance goals

Page 20

As with the other approaches discussed, one limitation is that these reports and recommendations do not exist for all analytes. Virtually none of these protocols, studies or reports make any allowances for lab to lab or method to method differences in results. None suggest interpretation of results using lab specific reference intervals. This means that there is increasing pressure on manufacturers and laboratories to minimize or eliminate these differences. As we all know this is not simple task for a number of analytes, but progress is being made and will continue to be made. Finally, these recommendations are clinically based and focus on what is desirable clinically. There have been a couple of cases where the performance recommendation cannot be met by any method in current use. Technology has not caught up with the perceived medical need.

SIEMENS

Determining Total Allowable Error

Recommended Hierarchy for Specifications:

1. Clinical outcome studies
2. Clinical expert opinion
3. Biologic variation
4. Professional recommendations
5. **Regulatory requirements**
6. State of the art

1999 Stockholm Conference:

Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Strategies to set global analytical quality specifications in laboratory medicine. Consensus agreement. *Scand J Clin Lab Invest.* 1999;59:585

Page 21

Next in the hierarchy are the recommendation made by regulatory agencies.

Total Allowable Error:
Regulatory Requirements

SIEMENS

National and state regulatory agencies have established acceptable limits for EQA/PT performance

CMS Centers for Medicare & Medicaid Services

Wadsworth Center
New York State Department of Health

RCPA Quality Assurance Programs Pty Limited

BUNDESÄRZTEKAMMER

OMP-LS
ONTARIO MEDICAL ASSOCIATION

Page 22

Agencies in many countries and even state agencies here in the US manage External Quality Assessment (EQA) or Proficiency Testing (PT) programs and have established acceptable performance limits for these inter-laboratory testing programs. If these limits are used to establish Total Allowable Error, we can then set as a goal detecting any change in method performance that would cause a failure with an EQA or PT result.

This approach to establishing Total Allowable Error has been very popular in most of the literature articles about Total Allowable Error and these limits are often listed in tables in these articles and in some software as recommended values for TEa. Let's look at some examples.

Total Allowable Error:
Regulatory Requirements - Examples

SIEMENS

Example:

CLIA '88 performance goals for proficiency testing

- Often used as examples in literature and software for Total Allowable Error limits
- Less than half the typical laboratory menu of analytes has CLIA PT goals
- Goals created by committee consensus based on 1980's technology
- **Useful resource – not a gold standard**

Total Allowable Error based on CLIA PT limits	
Glucose	120 mg/dl ± 10%
Sodium	115 mmol/L ± 3.47%
PSA	None
TSH	4.0 µIU/ml ±21%

Page 23

In the US the CLIA regulations have established performance criteria for a number of analytes.

Here are the CLIA goals for our example analytes With our example, we can find CLIA goals for Glucose, Sodium, and TSH and we can use the goals to set Total Allowable Error specifications at our chosen medical decision points. However there is no CLIA performance goal for PSA as is the case for many analytes and most immunoassays.

These performance goals can be used as the Total Allowable Error goal. However, these CLIA performance goals were established prior to 1992 using a consensus process and are based on the expected performance of analytical systems in use at that time. They don't reflect well current performance or necessarily medical needs and, most importantly, the goals are only set for about 40 analytes. These goals can be a good resource when establishing Total Allowable Error, but they are not a gold standard.

Total Allowable Error:
Regulatory Requirements

SIEMENS

Challenges:

- Acceptable limits not defined for all analytes
- While limits may be based on clinical requirements, may be altered to meet practical needs of PT/EQA programs
- Limits must incorporate allowances for factors such as sample stability, capabilities of older technology, matrix interactions

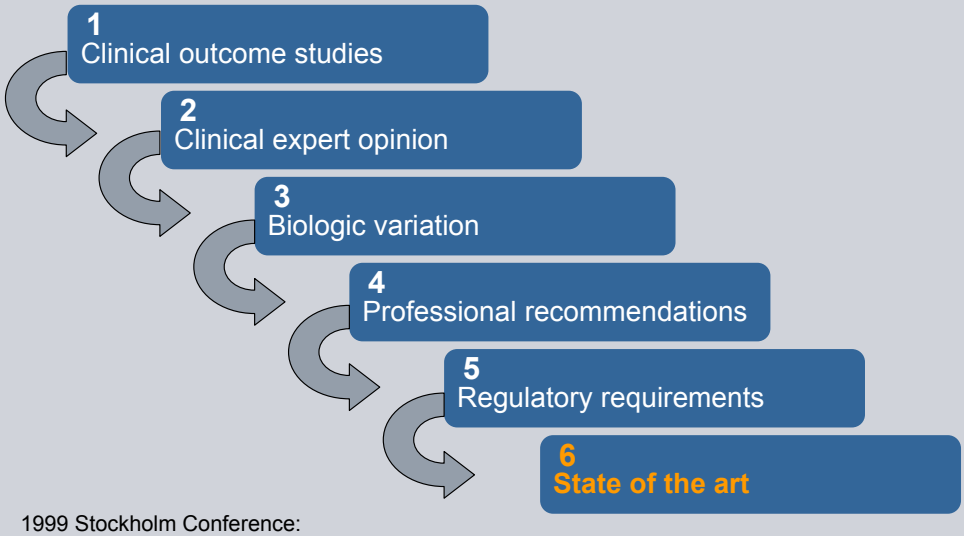
Page 24

There are some challenges to using EQA or PT limits for Total Allowable Error. Especially in the US many analytes commonly part of the labs menu do not have CLIA defined limits. Further, while these limits are often based on medical usefulness criteria, the actual limits are modified to meet the needs of the EQA / PT program. Things like sample stability, possible matrix interactions, the need to cover a wide range of analytical technology, etc. often drive adjustment of the medically derived limits to meet the practical constraints of an EQA / PT program.

SIEMENS

Determining Total Allowable Error

Recommended Hierarchy for Specifications:



1999 Stockholm Conference:

Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Strategies to set global analytical quality specifications in laboratory medicine. Consensus agreement. *Scand J Clin Lab Invest.* 1999;59:585

Page 25

Finally we have performance goals based on the current performance of the available methods. This is the final default criteria if nothing better can be found. This should be our last resort, not our first choice.

SIEMENS

Establishing Total Allowable Error

- No one approach can "do it all"
- There is no simple single formula to set TEa
- Tables in software and literature are examples, suggestions – **not** standards
- Goals are driven by medical need, clinical input is important

Determining TEa is not simple

Page 26

Setting Total Allowable Error goals for all analytes is the hardest part of developing an efficient and effective QC protocol. There is no simple, one right way to estimate Total Allowable Error. The example tables from literature articles are just that, examples. They are not standards or necessarily the best approach for us to use. We need to keep in mind that the goals are primarily clinical in nature.

SIEMENS

Establishing Total Allowable Error

- Select approach for each analyte
- Develop goals in collaboration with clinical customers
- Validate goals against analytical capability – is the goal practical ?

Determining TEa is not simple → Need to use reasoned judgement

We need to use a combination of approaches and work in collaboration with our clinical colleagues to establish our allowable error goals. Then, once we have a proposed set of goals, we need to validate them against the capability of our instruments and methods. It does no good to set a performance goal that no instrument or method can achieve. We always have to balance what we would like against what is possible.

SIEMENS

Establishing Total Allowable Error

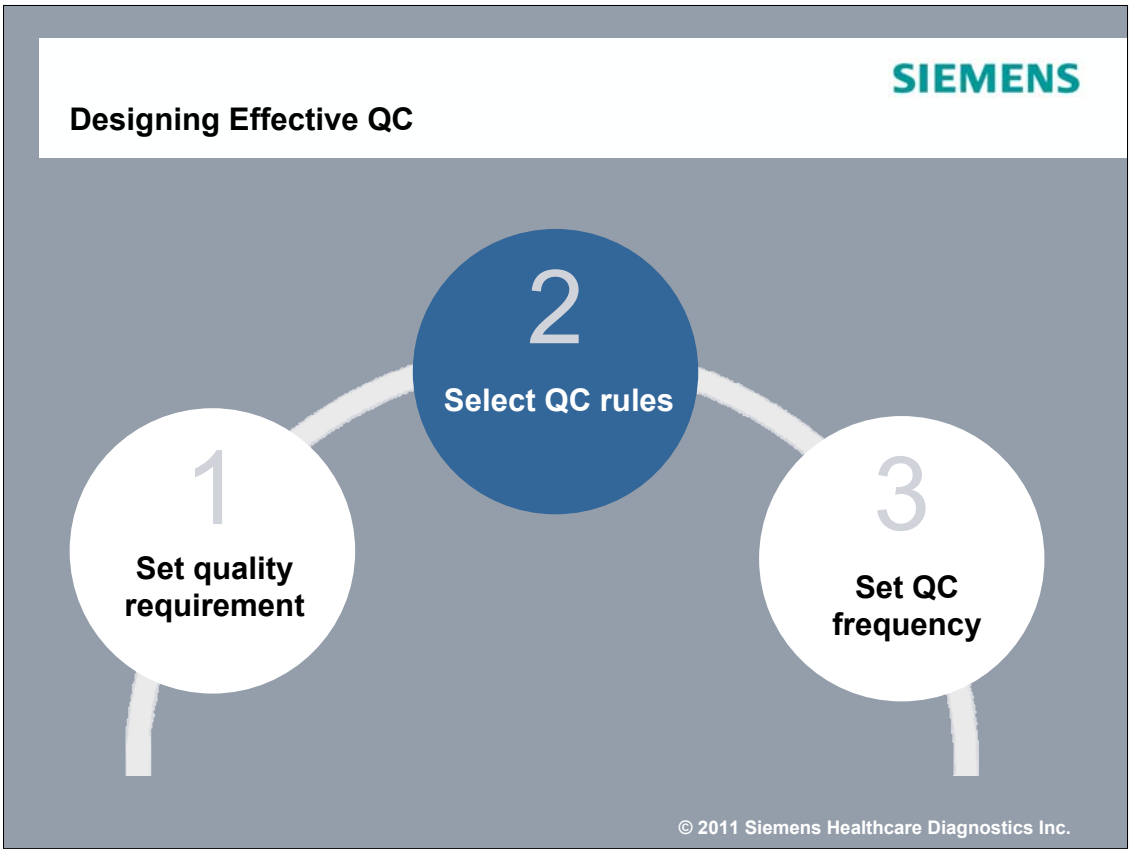
- Established TEa values can be used to help select methods
- Clinically based TEa values remain consistent unless clinical need changes

Determining TEa is not simple → Need to use reasoned judgement → TEa can be used with different methods

Page 28

Establishing total allowable error goals for all analytes in the lab takes a fair amount of time and effort and is never easy. However, once it is done. It is essentially done for all time. Since the Total Allowable Error goals are not based on how current methods perform, but rather on how results are used, once the goals are agreed on they can be used for a long time with different instrument systems. So, in the long run, the effort to set these goals is worth it.

Once we have established our Total allowable Error, we can use it to help select the optimal QC rules for our analytes ...



Once we have established our quality requirement, we can use this to select our QC rules. However, before we discuss how to use the quality requirement in rule selection, we need to review some key concepts behind the function of QC rules so the selection process makes sense.

SIEMENS

Types of Rules

Failure Rule:

If QC results fail this rule, method is considered "out of control"

Warning Rule:

If QC results fail this rule, considered an early warning alert.

Method is still "in control" and results are still reported without delay. Failure of a warning rule suggests there may be something worth investigating **while results are still reported**

Challenge:

Warning rule concept can be confusing

Warning rules are often treated as failures – testing is halted, results not reported. Negates the point of a warning rule

Page 30

An initial step in selecting rules is to decide which type of rules we want to use. There are two basic types. Failure rules are designed so that, if the QC data fails the rule, we say the method is out of control and we stop reporting results until further action is taken. Clearly this is the most common type of QC rule and all QC protocols need to be based on one or more failure rules.

We also have warning rules. These are rules that typically have too high a false positive rate to be effective failure rules, but they can function very well to give us an early indication that some change in performance may be occurring and allow time to investigate before we trip the failure rule. With a warning rule, if the QC results fail the rule, WE do NOT stop releasing results. Instead, we recognize that the method is still acceptable, but something may be happening. So we start to investigate to see if there really is an issue without interrupting the work flow.

The usual problem with using warning rules is that pretty soon everyone starts treating them as failure rules and stops reporting when the warning rule trips. This negates the whole idea of a warning rule and makes the QC process very inefficient.

SIEMENS

What rules are used?

Options include:

- Single rule
- Multi-rule

Most Commonly Used

- Mean & Range
- Cumulative sum (Cusum)
- Weighted moving averages
- Using patient results
- Others

So what rules are available to us to use. There are actually quite a lot of options. Single rule protocols and multi-rule protocols are the most commonly used and we will discuss them in some detail.

SIEMENS

What rules are used?

- Mean & Range
- Cumulative sum (Cusum)
- Weighted moving averages
- Using patient results
- Others

Not very widely used:

- All rely on calculations made using each QC result
- Have never been popular for manual application
- Computers can readily do the math
- Not commonly available on instruments or LIS
- Can be quite effective

These other options: Mean & Range, cumulative sum, weighted moving averages, using patient results and others like multi-variate approaches are all well documented in the literature and can be very effective and efficient. They have not been widely used because they pretty much all require that calculations be made each time a QC result is evaluated. In the past this was not practical for many labs. However, now all the instruments have powerful computers, many labs use middleware products that use powerful computers and most all labs are connected to LIS systems that can perform the calculations. However, if we look at the QC support software on our instruments, our middleware, and our LIS systems, we don't find these options available. So we are still left with single rule and multi-rule protocols as the most practical because they are well supported.

SIEMENS

What rules are used?

Single rule

Multi-rule

- Currently only readily available rules
- Only rules typically supported in software

Page 33

For now we will focus on Single rule protocols and multi-rule protocols since these are the most readily available procotols and the only ones generally supported by the software we use to manage QC.

SIEMENS

QC choices - Single rule

Single rule

Multi-rule

- One rule applied to each QC result
- Simple: If the rule fails, method is “out of control”

+/- 2 SD

- Historically, very commonly used
- Very inefficient rule
- False reject rate too high
- Can only be effectively used as a warning rule

A single rule protocol is just what it sounds like. A single QC rule is applies to each QC result as it is generated. If the result fails the rule, the method is deemed “out of control”.

Historically, the single rule +/- 2 SD has been the most commonly used. This is a very inefficient rule due to it’s very high false positive failure rate, especially when used with multi-level QC material for for many methods concurrently. It can be an effective warning rule … but most of the time the warning rule is actually used as a failure rule and we have gained nothing.

SIEMENS

QC choices - Single rule

Single rule

Multi-rule

- One rule applied to each QC result
- Simple: If the rule fails, method is “out of control”

Other rules can be used

- Reduce false rejection rate
- SD multiplier can be:
  2.5, 2.58, 3.0, 3.5
- Optimal rule balances low false reject rate and method performance

However, +/- 2 SD is not the only possible single rule. SD multipliers like 2.5, 2.58, 3 and even 3.5 can be effectively used to control the false positive rate and reliably detect change in performance. Notice that it is not required by statistics or science that the multiplier of the SD used for a single rule needs to be a whole number. The only reason most rules used historically have been whole numbers is that those rules were developed when we were doing the math in our heads.... And whole numbers are easier to work with. Today with computers doing the math, the multiplier can be any value we want in order to get the detection or false positive rate we desire. A multiplier of 2.58 SD gives us a false positive rate of exactly 1% per method per control. What is critical is to match the choice of rule to the quality requirement and the usual performance of the method. We will look at this in detail in a bit.

SIEMENS

QC choices - Multi-rule QC protocols

Single rule

Multi-rule

- Series of rules to validate QC results
- If one or more rules fail – method is “out of control”
- Each rule alone may not be ideal; taken together they provide effective QC monitoring
- Rules designed to assist in detecting trends
- Details are important: rules must be used exactly as designed

The other commonly available choice for QC rules is a multi-rule protocol. As the name implies, multi-rule protocols use a series of several rules to evaluate QC results. If the QC result fails one or more of the rules, the method is deemed “out of control”. The individual rules used are selected to have very low false positive rates. As a consequence, they often focus on specific types of errors and, each used alone may not be completely effective in detecting all changes in performance . However, used together, they reliably detect changes with a low overall false positive rate and ... looking at which of the rules failed can often provide useful information on possible root cause. Key to using these rules is understanding that each rule must be applied in a very specific way to be effective. We have to pay attention to these specific requirements or we may lose the overall effectiveness of the process. Details are important.

SIEMENS

QC choices - Multi-rule Example

Single rule

Multi-rule

Best known: "Westgard Rules"

- Rules taken from statistical process control used by other industries
- Chosen to keep false positive rate to 5%
- Published set of rules is a "toolbox"
- Other multi-rule approaches can be used

CLIN. CHEM. 27/3, 493–501 (1981)

A Multi-Rule Shewhart Chart for Quality Control in Clinical Chemistry

Submitters: James O. Westgard, Patricia L. Barry, and Marian R. Hunt, *Depts. of Medicine, Pathology, and Clinical Laboratories, University of Wisconsin, Center for Health Sciences, Madison, WI 53792*
Torgny Groth, *Group for Biomedical Informatics, Uppsala University Data Center, S-750 02 Uppsala, Sweden*
Reviewers: Robert W. Burnett, *Clinical Chemistry Laboratory, Hartford Hospital, Hartford, CT 06115*

In recommending the multi-rule Shewhart procedure, the objectives have been to provide (*a*) simple data analysis and display via control charts, such that computerized data handling is not necessary; (*b*) easy adaptation and integration into the existing control practices in clinical laboratories; (*c*) a low level of false rejections or false alarms; (*d*) an improved capability for detecting analytical errors; and (*e*) some indication of the type of analytical error occurring when a run is rejected, to aid in problem solving.

Principles

The best example of a multi-rule protocol is also the most widely known, the so called "Westgard Rules". Dr. Westgard and three other authors published the paper introducing these rules 30 years ago. The rules used were selected from statistical control rules used in other industries. Dr. Westgard selected rules that would best fit the way a clinical laboratory operates and which would have a very low false positive rate. The rules as described in the riginal article and in all subsequent writings are not a fixed set of required rules, but rather a tool box of rules that can be used. There are other multi-rule protocols available, but they all essentially work the same way. Let's look at Dr. Westgard's proposed rules in a little more detail.

SIEMENS

Example: "Westgard" Rule Toolbox

Scatter	N = 2 or 4	Bias	N = 3 or 6
1_{3s}	2_{2s}		$2 \text{ of } 3_{2s}$
R_{4s}	4_{1s}		3_{1s}
	8_x		6_x
		7_t	

2_{2s} means …

2 consecutive QC results that both exceed **2 SD** in the same direction

N = number of QC samples per run

Page 38

Here is Dr. Westgard's rule tool box. As you can see, some rules are designed to detect increased scatter or imprecision. Other rules are designed to detects changes in bias or shifts. We can also see that which rules you should use depends in part on how many QC results are being evaluated together. We have one set of rules for when we use 2 levels of controls and a somewhat different set if we use three levels of controls.

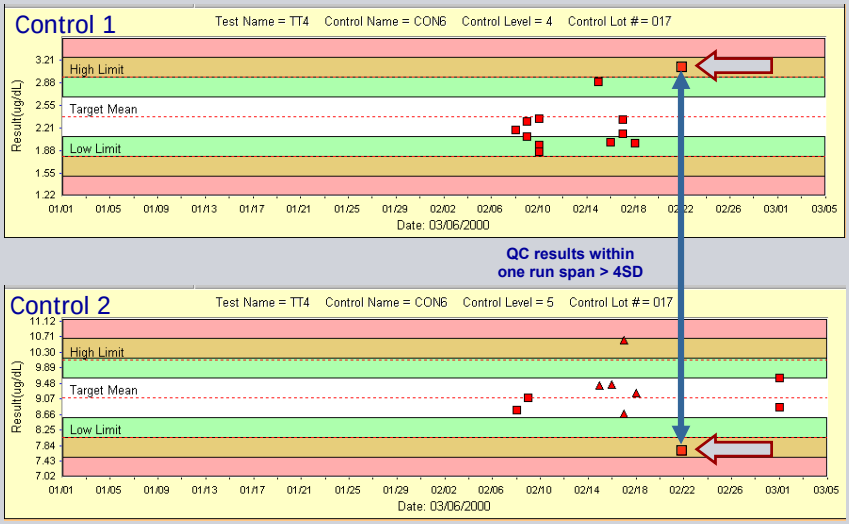
The notation may seem strage at first, but it is easy to understand. 2 2s means two consecutive QC results that both exceed 2 SD on the same side of the mean. Similarly 4 1s would mean 4 consecutive QC results all exceeding 1 SD on the same side of the mean. Let's look at these other rules

SIEMENS

R_{4s} Examples

Range

Example 1:



Page 39

The R 4 s rule looks at the range spanned by two controls within the same run. If the span exceeds 4 SD, then the rule fails. Note this applies only to controls run together in a single run and that they do not have to be consecutive. If we are using three levels of control, if any two of the three results show a span exceeding 4 SD ,then the rule fails.

SIEMENS

N_x rules (6_x, 8_x, 9_x, 10_x, 12_x)

N consecutive results on the same side of the mean

- **Very** sensitive rules to detect shifts in mean

For many methods, may be too sensitive

Can be used as warning rule or not used at all

Use requires careful setting of target mean and frequent checks to see if update to mean is needed

Many laboratory professionals like to use the 10_x and 4_{1s} and similar rules as "warning rules," using those trends and shifts as a way to get an early eye on a problem, even if QC design doesn't mandate those rules. That's fine, but if it starts to make you chase ghosts in the method, it's counter-productive."

Westgard, J.O., *Ten Ways to do the Wrong QC Wrong*, Westgard QC Inc., 2007

Page 40

The "N x" rules are interesting. The basis of the rule is "N" consecutive QC results all on the same side of the mean. Values used for "N" have been 6, 8, 9, 10, 12. These rules are designed to detect changes in bias or shifts and they are very sensitive … sometime too sensitive. Dr. Westgard has recommended using this type of rule as a warning rule in most cases or not using it at all. These rules are best saved for methods where there is little room for change in method performance. That is a very small minority of methods as we shall see.

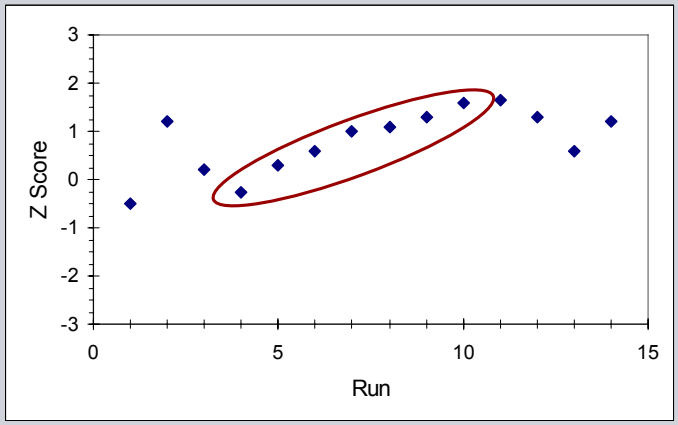
If use of these rules is contemplated, it is absolutely critical that the target mean be carefully set using data from the instrument and that the mean be checked and updated regularly. One of the fastest routes to frustration and highly inefficient QC is to try to use these rules with QC targets taken from a package insert or IFU. That will virtually never work because the actual instrument mean almost never matches the IFU mean exactly. This is normal and expected, but if the IFU mean is used as a QC target mean, these "Nx" rules will consistently fail because of that difference.

SIEMENS

Trend rule: 7_t

7 consecutive results, each one greater (less) than the preceding result

Popular in Europe



The last rule we will look at is the 7T. This is a trend rule that has been popular in Europe. It requires that 7 consecutive QC results each be greater than (or less than) the result immediately before. This is not the same as the "N x" rules since they only require that the results be on the same side of the mean. Here each result must be numerically greater than, or less than the one before it.

SIEMENS

"Westgard" Multi-rule definitions:

Run

- Number of controls evaluated at one time for a particular assay
  - Can be multiple levels of one control material or several completely different control materials

Within

- Comparison of control results **within the same control sample** (level) across multiple runs
  - Example: last 3 results for level 2 for Glucose
    Results will be from different runs and can be from different days

Across

- Comparison of control results **across different control samples** within the same run
  - Example: current results for levels 1,2 & 3 for TSH
    Can be different control samples (levels)

Page 42

There are also a couple of concepts or definitions that are important to effectively using the Westgard Rules. The first concept is the "run". The term comes from the days when patient samples used to be tested in separate defined batches or runs, rather than continuously. As applied to these QC rules, the concept of run really iss about how many QC results will be evaluated together at one time. If we use a bi-level QC material and run both levels together, then the run is 2 QC samples and the rules are applied to both results simultaneously once both results are available.

The other two concepts are "within" and "across". These terms indicate how the rules are applied to the QC results. As originally used by Dr. Westgard, "within" refers to applying the rules within a single control material, like BioRad level 1. This often means looking back to previous QC runs on other days to have enough data to apply the rule. "Across" implies applying the ruule across different control materials within a single QC run. This would be applying the rule to the two QC levels run just now.

SIEMENS

"Within" & "Across"

	Level 1		Level 2		Level 3	
Date	Result	Z	Result	Z	Result	Z
21-Aug	101.6	2.41	231.3	1.37	345.9	-0.33
20-Aug	86.0	0.12	202.6	-0.55	387.6	1.43
13-Aug	78.3	-1.01	204.3	-0.44	342.8	-0.46
9-Aug	81.9	-0.49	227.7	1.13	360.9	0.30
8-Aug	87.8	0.39	216.8	0.40	384.6	1.30
7-Aug	87.9	0.41	220.6	0.65	359.5	0.24
6-Aug	74.6	-1.55	256.2	3.04	394.4	1.72

Across controls
Within a run

Within a control
Across runs

Page 43

Here is another way to visualize the concepts of within and across. Most QC rules are designed to be applied both ways The idea behind looking back to previous days is to gain sensitivity to detect changes early on by using more data. This is really what we instinctively do when we look at the QC graph and review the data from previous days. Applying the rules this way just makes that look back part of the QC rules.

SIEMENS

Guidelines to Effective Use

1. Select the rules based on method performance

- Not required to use all the rules all the time
- Select rules that meet specific QC need
- Cannot trust random combinations
- Validated combinations are documented

Table 4. Summary of Control Procedures Appropriate for Different Numbers of Control Observations

No. control observations	Control rules for: Individual analytical runs	Control rules for: Consecutive analytical runs
1	1_{2s}	4_{1s}
2	$1_{3s}/2_{2s}/R_{4s}$	$4_{1s}/10_{\bar{x}}$
3	$1_{3s}/(2 \text{ of } 3)_{2s}/R_{4s}$	$9_{\bar{x}}$
4	$1_{3s}/2_{2s}/R_{4s}/4_{1s}$	$8_{\bar{x}}$
4–10	Mean/range	Trend analysis (*16*)
4–20	Mean/chi-square	Trend analysis (*16*)

Westgard et al, A Multi-Rule Shewart Chart for Quality Control in Clinical Chemistry, Clin Chem, 27, 493, 1981

Now let's look at some guidelines to the effective use of the Westgard Rules

First – Select the rules used based on method performance. We will discuss how to do this is detail in a few moments, but right now I want to make the point that … you are not required to use all the rules all the time. Even in the original paper that so many have referred to, Dr. Westgard selected the which subset of the rules to use based on the number of QC samples tested in each run. Today, the selection is driven by method performance. Key is that random combinations do not work. The rules have been validated to work in some very specific groupings. The specific groupings can be readily found on Dr. Westgard's website and even in the original paper as shown here.

SIEMENS

Guidelines to Effective Use

1. Select the rules based on method performance
2. Rules used to evaluate ALL the QC results from a "run" as a group

Protocol: QC run is defined as one replicate each of three levels of QC

Rules used: 1_{3s}, 2 of 3 $_{2s}$, R_{4s}

Level 1 result
Level 2 result
Level 3 result

With all three results available, can apply rules

Next, the rules are designed to be applied to QC results as a Run … not to each individual QC result as it is generated. This clearly becomes critical for a rule like 2 of 3 2S. If you don't have all three QC results, how can you apply the rule. This has rarely been an issue when people were manually applying the rules, but it can be an issue with computerized applications.

SIEMENS

Guidelines to Effective Use

1. Select the rules based on method performance
2. Rules used to evaluate ALL the QC results from a "run" as a group
3. Once a "run" fails, future "runs" are evaluated applying rules only results obtained after the rejected "run"

Once we have a rule failure, the data used to evaluate the rules cannot come from prior to the rule failure. Let's see how this works

SIEMENS

Reset rules after rejected run

	Level 1		Level 2		Level 3	
Date	Result	Z	Result	Z	Result	Z
28-Aug	91.2	0.88	196.2	-0.98	346.4	-0.31
28-Aug	81.2	-0.59	224.2	0.89	357.0	0.14
27-Aug	96.2	1.61	229.3	1.24	342.4	-0.48
24-Aug	83.7	-0.22	216.2	0.36	350.9	-0.12
23-Aug	81.1	-0.59	207.5	-0.23	340.2	-0.57
22-Aug	78.3	-1.00	221.3	0.70	376.2	0.95
22-Aug	99.8	2.13	212.4	0.10	357.9	0.18
21-Aug	101.6	2.41	231.3	1.37	345.9	-0.33
20-Aug	86.0	0.12	202.6	-0.55	387.6	1.43
13-Aug	78.3	-1.01	204.3	-0.44	342.8	-0.46
9-Aug	81.9	-0.49	227.7	1.13	360.9	0.30
8-Aug	87.8	0.39	216.8	0.40	384.6	1.30
7-Aug	87.9	0.41	220.6	0.65	359.5	0.24
6-Aug						

STOP

2 of 3_{2s}

Once we have a failed run, we start over with the data used for rules going forward. So it will be 4 runs into the future before we can apply the 4 1s rule within a single control. However, this only applies to the QC rules. When we use this data to calculate a mean or SD, we use all the data except from the specific run that had the problem.

SIEMENS

Reset rules after rejected run

	Level 1		Level 2		Level 3	
Date	Result	Z	Result	Z	Result	Z
28-Aug	91.2	0.88	196.2	-0.98	346.4	-0.31
28-Aug	81.2	-0.59	224.2	0.89	357.0	0.14
27-Aug	96.2	1.61	229.3	1.24	342.4	-0.48
24-Aug	83.7	-0.22	216.2	0.36	350.9	-0.12
23-Aug	81.1	-0.59	207.5	-0.23	340.2	-0.57
22-Aug	78.3	-1.00	221.3	0.70	376.2	0.95
22-Aug						
21-Aug						
20-Aug						
13-Aug						
9-Aug						
8-Aug						
7-Aug						
6-Aug						

Prior QC data not used to apply rules going forward

STOP

2 of 3_{2s}

Page 48

Once we have a failed run, we start over with the data used for rules going forward. So it will be 4 runs into the future before we can apply the 4 1s rule within a single control. However, this only applies to the QC rules. When we use this data to calculate a mean or SD, we use all the data except from the specific run that had the problem.

SIEMENS

Guidelines to Effective Use

1. Select the rules based on method performance
2. Rules used to evaluate ALL the QC results from a “run” as a group
3. Once a “run” fails, future “runs” are evaluated using only results obtained after the failed “run”
4. Rules were selected for manual application to QC run in a batch

Recognize that the rules were developed in the 1970’s and were designed to be realtively simple for people to use manually

SIEMENS

Manual vs. Computerized rules

Rules chosen to be simple for a person to apply

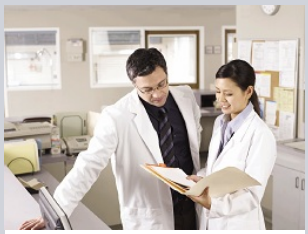
- Application fairly intuitive for people
- Can be applied by viewing plotted results

Page 50

It is not difficult to teach someone to manually look at graphed QC results and apply the Westgard rules. Keep in mind they were always meant to be evaluated looking at a QC graph. It was never intended that anyone would try to use the rules looking at columns of numbers on a page.

SIEMENS

Manual vs. Computerized rules



Critically review computerized implementations

- Not all rules may be implemented
- All rules may not function as described for manual application
  - Evaluating the QC "run" rather than each QC result as generated
  - Applying rules "within" and "across"
  - Using data from "failed" runs for rule evaluation
- Need to know exactly how implementation works
- Can be valid and useful, but must know how they work

Page 51

However, now most folks use some sort of a computerized implementation of the rules and there's the challenge. Most computer implementation of the Westgard Rules do not use the rules the way Dr. Westgard originally intended. Frequently not all the rules are available, especially those for three levels of control. Then the rules are often not applied "within" and "across" and finally the rules are often applied to each individual QC result as it is generated rather than collectively to the run.

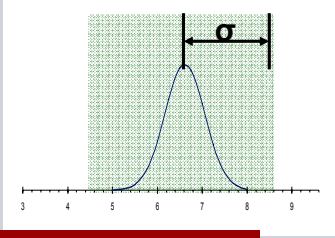
These differences do not mean that these implementations of the rules are not good and do not work. They can be effective and do the job, but it is important that we know exactly how they work and not assume that just because they are called Westgard Rules, they are exactly as described in the original paper.

SIEMENS

TEa and QC

Compare TEa to current method performance

- TEa sets the clinical error limit
- Method performance determines when change becomes significant



If typical method error is close to total allowable error, it will be very difficult to control assay performance to prevent exceeding the TEa

If typical method error is much less than total allowable error, it will be relatively easy to detect change in the assay's performance before exceeding the TEa.

The ratio of the method's typical error relative to the Total Allowable Error goal has been called the Sigma Metric

Page 52

Now, finally let's bring it all together and use our Total Allowable Error based quality requirement and our understanding of the QC rules to see how we can select effective and efficient QC rules for our methods

To do this we compare our TEa goals to the actual performance of our methods on the instrument we are using. This is where we make the connection between TEa goals and actual method performance.

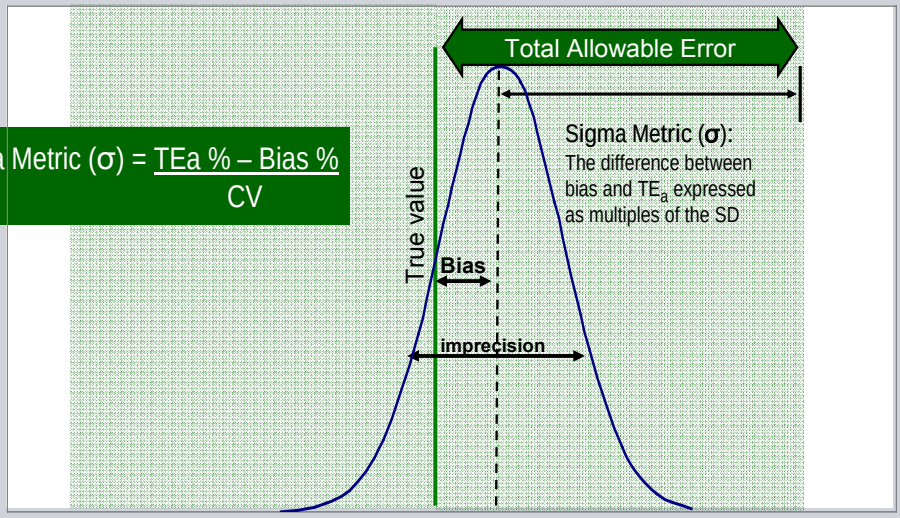
So, if Total Allowable Error is close to the actual performance of the assay, it may be difficult to monitor the assay and control it to prevent change in assay performance from impacting assay interpretation. However, if the actual method variability is small compared to the performance goal it will be easy to detect change in performance before it has an impact on patient care.

Recently, folks have begun taking the ratio of TEa to the method's variability as a guide to selecting QC rules. This ratio is called the Sigma Metric.

SIEMENS

What's a Sigma Metric ?

How much change in the analytical process can be tolerated



$$\text{Sigma Metric } (\sigma) = \frac{\text{TEa \% – Bias \%}}{\text{CV}}$$

Page 53

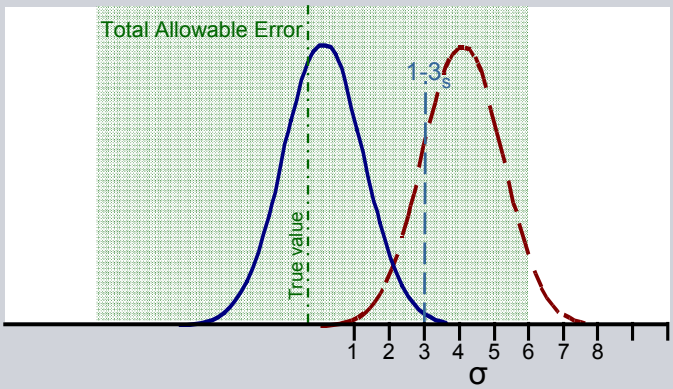
The Sigma metric is a measure of the difference between the actual method error and the Total Allowable Error. Here we see the performance of an assay relative to the "true" value and the Total Allowable Error. The Sigma Metric is calculated by subtracting the assay's bias from the Total Allowable Error goal and then dividing that difference by the assay CV. This gives the difference between current assay performance and the error goal as multiples of the CV or SD.

As you might expect, the ideal is for the Sigma Metric to be 6 or higher. Let's see how we can use this value to determine what QC rules will be effective.

SIEMENS

High Sigma Methods and QC Rules

Simple single rule QC will reliably detect method change before TEa is reached



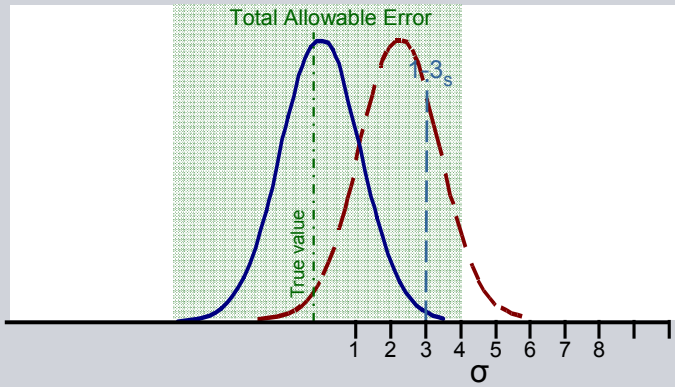
Page 54

With high sigma methods, the difference between typical performance and the total allowable error limit is sufficiently large that a simple single rule protocol like +/- 3 SD can readily catch any significant change in method performance before we exceed the allowable limit and still have a very low false positive rate.

SIEMENS

Low Sigma Methods and QC Rules

More complex multi-rule QC protocols may be needed



Page 55

On the other hand, a low sigma method doesn't have the same cushion to work with. In this case using +/- 3 SD will not be effective because we will have exceeded the error limit well before a 3 SD limit will consistently indicate the change in performance. In this case a multi-rule protocol will be more effective and which rules to use will depend on the sigma.

SIEMENS

Selecting QC rules

3 4 5 6

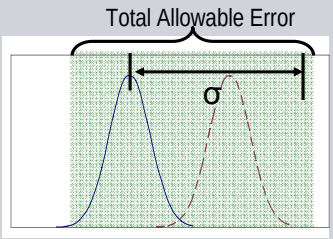
Sigma Metric

QC Rules

$1_{3s}/R_{4s}/2_{2s}/4_{1s}/8_x$	$1_{3s}/R_{4s}/2_{2s}/4_{1s}$	$1_{2.5s}$	$1_{2.5s}$	1_{3s}	$1_{3.5s}$
n=6	n=4	n=4	n=2	n=2	n=2

σ metric will be different for each method:

What does that suggest about the rules used ?



JO Westgard, Six Sigma Quality Design & Control, 2nd Ed., Westgard QC inc., 2006

Page 56

When we use the sigma metric to help select QC rules we find there is a continuum of which QC rules work best at which sigma metric.

If the assay's sigma metric is 5 or greater, it becomes fairly easy to detect change in performance before the analytical performance can impact decision making and the QC protocol used can be very simple.

If the assay's sigma metric is between 4 and 5- it's still fairly easy to catch change, but slightly more powerful QC rules are needed

If the assay's sigma metric is between 3 and 4 – it is more difficult to catch performance changes before they impact decision making, but it is still practical with reasonable QC protocols. The closer we are to 3 sigma the more complex the rule set.

If the Sigma metric is less than 3, we need all the QC rule support we can get and even that may not be able to effectively monitor changes in the assay's performance to prevent any impact on decision making using statistical QC protocols alone.

Fortunately, most current assays fall into the 4 or better sigma range and so are OK. However, in the menu of almost every system are a few that do not. If that's the case, and an alternate better method is not practical, then we have to use maximum statistical QC and know that even that may not detect all significant changes.

Since these choices are based on Sigma, it seems to suggest we could have multiple QC protocols in the lab

SIEMENS

Applying the Rules

Method	σ	+/- 3 SD n=2	+/- 2.5 SD n=2	+/- 2.5 SD n=4	Multi-rule n=6
Glucose	4.8				
Creatinine	7.5				
BUN	3.3				
K+	5.0				
Na+	2.9				
Calcium	4.5				
LD	6.2				
CK	9.5				
CEA	4.0				
Cortisol	6.2				
Estradiol	3.4				
Folate	6.9				
Microalbumin	9.2				
PSA	6.1				

Page 57

When you do a sigma analysis and look at the results, it's easy to see that we will certainly not use the same QC protocol for everything in the lab and probably not even for all the methods on a single instrument. How are we supposed to manage that ?!

SIEMENS

Applying the Rules

+/- 3 SD n=2	+/- 2.5 SD n=2	+/- 2.5 SD n=4	Multi-rule n=6
Creatinine	Glucose	Calcium	BUN
LD	K+	CEA	Na+
CK			Estradiol
Cortisol			
Folate			
Microalbumin			
PSA			

How does this work ?

- Currently supported in software /one time configuration
- Could test one QC panel of 2 levels for all; 2nd panel for 5 methods
- At the bench, no difference ⇨ QC is run – did rule fail ?

Page 58

As we work it through we can see that methods get grouped into one of three or four different QC protocols based on their sigma value. So we only have a small number of different QC protocols. Still who can remember this ? No one can or needs to. The QC software on most instruments today allows QC rules to be assigned on a method by method basis. A number of Siemens systems have supported this for more than 10 years. So we don't have to remember, the computer does. We configure the QC software one time and it remembers from that point on. Then we can use QC panels to easily schedule the number of QC samples appropriate to each method. So that, looking at QC on a daily basis, nothing changes, the QC software flags results that fail the rules and we follow up.... Regardless of the QC protocol.

SIEMENS

Practical Challenges

To calculate need:

$$\text{Sigma Metric } (\sigma) = \frac{\text{TEa \%} - \text{Bias \%}}{\text{CV}}$$

CV: easily obtained from QC data. Just be sure to use enough data over enough time to accurately reflect method

TEa: already discussed challenges with determining TEa

Page 59

As is often the case when we try to take a good idea and use it in the real world, there are some practical challenges. To estimate the Sigma metric we need three values: Total Allowable Error, bias and the CV.

CV is relatively straight forward if we have QC samples that are targeted near the decision points of interest. We can use the CV from the QC material. We have to make sure that we are calculating the CV using enough data. 10 values is no where near enough and even 20 values will not give a robust estimate of CV. It is best to use data from several months of QC testing if possible.

We have already discussed the challenges with determining total allowable error so won't go over that again. However we recognize there is effort involved is choosing the best value to use.

SIEMENS

Practical Challenges

To calculate need:

$$\text{Sigma Metric } (\sigma) = \frac{\text{TEa \%} - \text{Bias \%}}{\text{CV}}$$

Bias: how to determine ? Bias compared to what ?

- QC or PT Peer group mean – often used, but this is most common result, not necessarily most accurate
- EQA/PT all method mean – in many cases it's only a peer group mean
- Result from "Reference Lab" – not usually reference method
- Reference method result for same sample(s) – best by far, but who has access to these methods ? Some PT target values
- One pragmatic approach: assume bias is zero

Page 60

Finally there is bias. This can be a difficult challenge. Bias represents how much our results differ from the true result on the average. But what is truth ? How do we know what the true value is ? In articles about using sigma metric, it is often suggested that we use the QC or PT peer group mean as our measure of truth. But is that really the best choice? The peer group mean is not necessarily the most accurate value only the most popular one. It is entirely possible that the peer group is generally more biased than we are.

In the past folks have used the all method mean from PT results as truth, and at one time it may have given a reasonable estimate. However, today for many, many methods there is a predominant market leader that most labs are using and the all method mean is really nothing more than the peer group mean for that method. If that method is unbiased, then it is fine ... but how do we know that method is unbiased?

We can send samples to a reference or commercial lab to get comparative results. However, often these labs use the same methods we do. Sometimes however, these large labs do have reference methods, or something very close, available. If that's the case then those results could give us a good estimate of bias. What we really want is comparative results for fresh patient samples from a real reference method. Unfortunately that is almost impossible to find. Reference methods are very manual and are usually not practical for routine use. So we cannot afford to set them up and often cannot find a lab that can. In recent times some PT programs have begun assigning target values using reference type methods and grading is against the reference result rather than the peer group. If that's the case, those TP targets may be useful.

One pragmatic way to get started using the concept of sigma metric even if we cannot find a good way to estimate bias is to assume bias is zero. If we do this, we can estimate a sigma metric and use it to help set up out QC and generally we will get close to the ideal. Most methods do not have large biases so this can work at a very basic level to help us get started. Then once we find an estimate of bias that we feel accurately represents method bias with patient samples, we can revise our estimate of Sigma metric and adjust accordingly.

SIEMENS

Estimating the Sigma Metric

Analyte	Bias	CV	Medical TE_a	CLIA TE_a	Biologic TE_a
Glucose	1%	2.3%		10%	6.9%
Na^+	0%	1.0%		3.47%	0.9%
PSA	N/A	5.0%		None	33.6%
TSH	1%	4.9%	20%	21%	22.8%

$$\text{Sigma Metric } (\sigma) = \frac{TE_a\ \% - \text{Bias } \%}{CV}$$

TSH Medical $\sigma = \frac{20\ \% - 1\ \%}{4.9\%} = 3.9$
(at 4.0 mIU/L)

TSH CLIA $\sigma = \frac{21\ \% - 1\ \%}{4.9\%} = 4.1$
(at 4.0 mIU/L)

TSH Biologic $\sigma = \frac{22.8\% - 1\ \%}{4.9\%} = 4.4$
(at 4.0 mIU/L)

Page 61

Looking at our example assays, there is only one, TSH, for which we have documented Error goals based on all three approaches medical use, CLIA limits and biologic data. Let's follow TSH through the process.

For the goal based on medical use we get a sigma metric of 3.9. Using the CLIA based goal we get a sigma metric of 4.1 and using the biologic goal we get a sigma metric of 4.4. All pretty much the same and all indicate that we can monitor and control TSH to meets these goals using standard statistical QC protocols.

However, that is not the case for all assays

SIEMENS

Analyte	Bias	CV	CLIA TE_a	Biologic TE_a	CLIA σ	Biologic σ
Glucose	1%	2.3%	10%	6.9%	3.9	2.6
Na^+	0%	1.0%	3.47%	0.9%	3.5	0.8
PSA	N/A	5.0%	None	33.6%	None	6.7
TSH	1%	4.9%	21%	22.8%	4.1	4.4

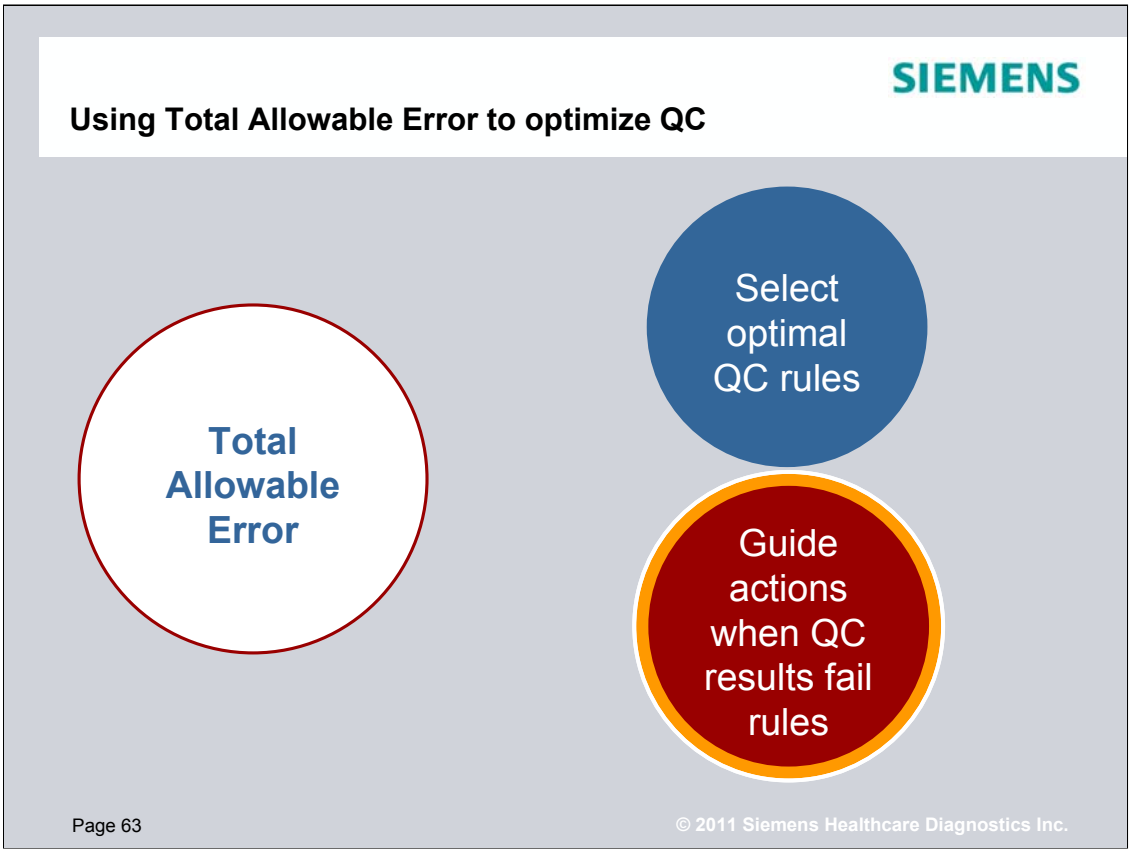
Challenges:

- For some, no method in routine use has performance to meet biologically based goal
- For others, no medical or CLIA based performance goals are available
- There is no simple uniform way to set goals

Page 62

When we look at our four example assays we see some of the challenges we face. For some assays the biologically based goals may not be achievable with current methods and technology. For other analytes, there may not be defined goals using criteria other than the biologic criteria. So we find that there is no simple uniform way to set Total Allowable Error goals and estimate the sigma metric. It becomes a decision based on available information and judgment.

However, it is worth the effort because it is so useful in helping us set up the most efficient and effective QC protocols.



There is another way that Total Allowable Error can help us in looking at QC results and that is to guide our actions when we have a QC rule failure.

SIEMENS

TEa and Actions after QC Rule Failure

If QC results fail the rule(s):

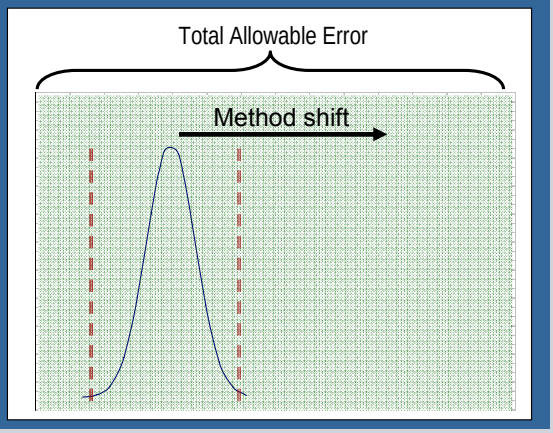
- If apparent change puts results near TEa limit – hold results, act now
- If apparent change is still well within TEa – can still report while investigating

TEa should *NOT* be used as QC limit for rules

Statistical QC detects change in performance

TEa allows the change to be put in context to determine appropriate follow up

Alternative to warning rules

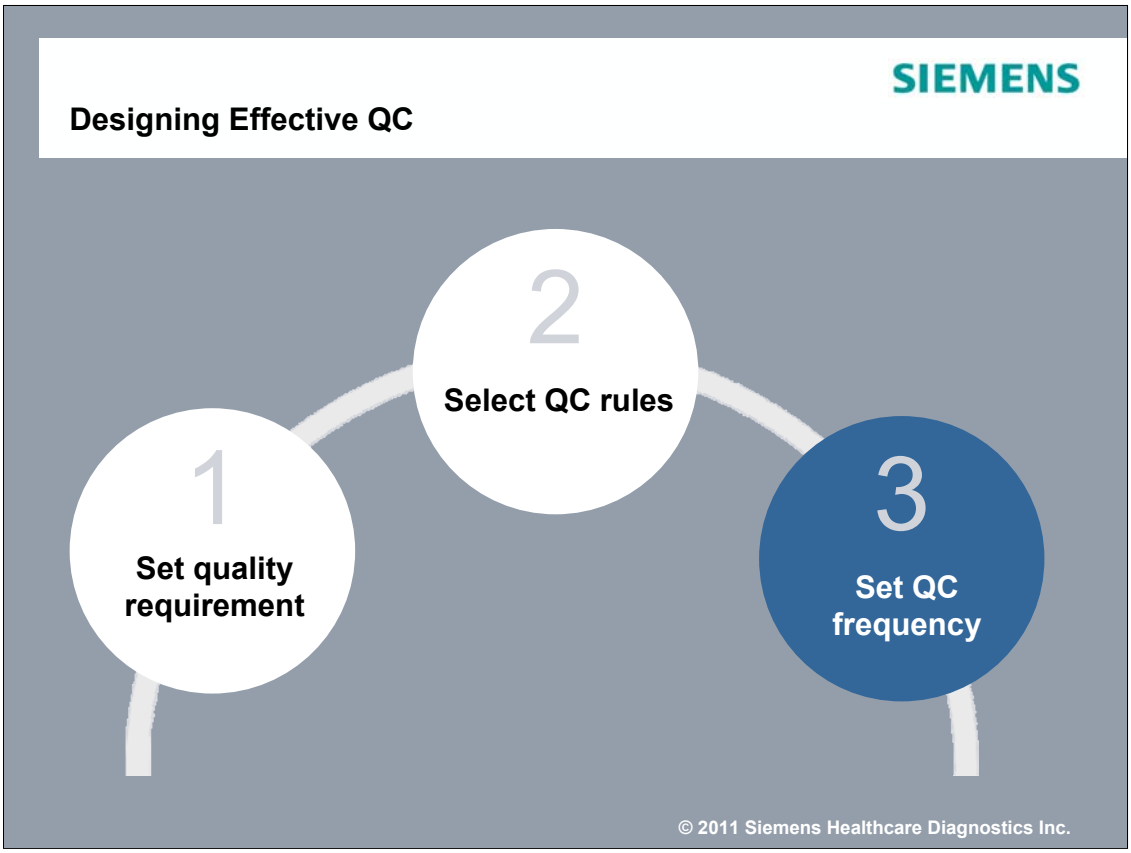


Page 64

Our QC rules are statistically based and are designed to detect any change in method performance. If the apparent change in performance puts assay results near the limit of the Total Allowable Error, then all results should be held until the investigation is complete and the issue resolved.

However, if the shift in performance cause a QC rule failure, but the results are still comfortably within the Total Allowable Error limit, then results can still be reported while the investigation is being done. This is because in spite of the change in method performance the error in the results still is not large enough to affect medical decisions.

Some points to note … This does NOT mean that we should use total allowable error limits as the acceptable limits for our QC rules. That would not work very well at all. We want our QC rules to work for us to detect any change in method performance. Then we can use total allowable error to put this change in context relative to medical decision making. Once we have put the method performance in context, we can make the appropriate decisions about how to proceed and whether patient results can be released. In this regard Total allowable error can function like a warning rule.



So we have set our quality requirement and used it to help select the optimal QC rules, now we need to establish when to test QC samples in order to finalize our QC protocol

SIEMENS

When do we test QC samples ?

Event based:

- Calibration
- New reagent lot
- Major maintenance
- Service

Will detect common reasons for change in method performance

Routine Monitoring:

- Testing for random error
- Random error is infrequent & unpredictable
- Spot check periodically

How often is periodically ?

Regulatory:

US – "two concentrations once each day of testing" unless you can use EQC; then it's effectively "No QC"

Germany – "twice within 24 hours, no more than 16 hours between events"

Page 66

When do we test QC samples? Generally there are two triggers for QC testing. One is event based. We test QC samples every time we do something that may have altered the performance of the system. Things like calibrate, maintenance, new reagent lots, etc. The second trigger is based on routine monitoring to detect random error. We know any analytical system can fail. We know these failures are random in nature and infrequent. So we cannot predict when they will occur. As a consequence we periodically test QC samples as a spot check for this random error. But how often is periodically?

Even regulatory agencies cannot agree. In the US CLIA says the MINIMUM is once every 24 hours of testing. In Germany, the requirement is twice in 24 hours with no more than 16 hours between events. So how do we decide ?

SIEMENS

Detecting random failure

Goal: detect failure before inaccurate patient results are reported

Based on risk to patients, not expected frequency of failure

Failure rate is very low

Occurrence cannot be predicted

Testing QC samples only checks single point in time

Test a QC sample with each patient sample !

WRONG !

Page 67

Our goal is to detect failure before any incorrect patient results are reported. So our goal is really risk based. We are more concerned about how many patient results might be incorrect than we are about how often the system might fail.

We know the failure rate is low. We cannot predict when the failure will occur. We know that testing QC samples can only tell us how the system is performing at the moment the QC sample is tested. So the obvious conclusion is to test a QC sample with every patient sample just to be sure ! WRONG !!! Clearly this conclusion is not workable. It is completely impractical because of the realities of workflow and the associated costs. So what do we do. We have to balance our need to reduce patient risk with the practical realities and costs. Let's look at cost in more detail

SIEMENS

Cost of QC

Direct cost:

- QC material
- Reagents
- Disposables
- Labor time

Easy to assess

Less QC = lower cost

Indirect cost:

- Delay in reporting
- Failure cost
  - Look back
  - Phone calls
  - Corrected reports
  - Incorrect treatment
  - Liability

Harder to quantify

More QC = lower cost

True Cost = Direct cost + Indirect cost

Page 68

The direct costs of QC are fairly easy to understand and estimate. They include the cost of the QC material, the reagents and disposables used and the labor cost. But there is another cost to QC – the indirect costs. The costs resulting from delayed reporting of patient results because we are running QC samples on the instrument and investigating all the false positive QC rule failures before we report results. There's also the failure cost. This is the costs associated with the occurrence of a QC rule failure that is then determined to be a true failure. The costs of any look backs at patient results. The direct costs of any repeat testing of patient samples. The cost of phone calls and corrected reports. The costs of incorrect treatment decisions because of incorrect labs results and the potential liability costs of the incorrect results. Fortunately these last two are not often a big concern because few treatment decisions are made solely on the basis of a single lab result. However, it can happen.

To understand the true cost of what ever QC protocol we use, we have to estimate the indirect costs and factor that into the total cost. Direct costs are easy to assess and generally, the less QC we do, the lower the direct cost. Indirect costs are tougher to estimate and generally the less often we test QC sample, the greater the potential indirect costs.

SIEMENS

Cost of QC

Indirect cost:

- Delay in reporting
- Failure cost
  - Look back
  - Phone calls
  - Corrected reports
  - Incorrect treatment
  - Liability

Estimating Failure Cost:

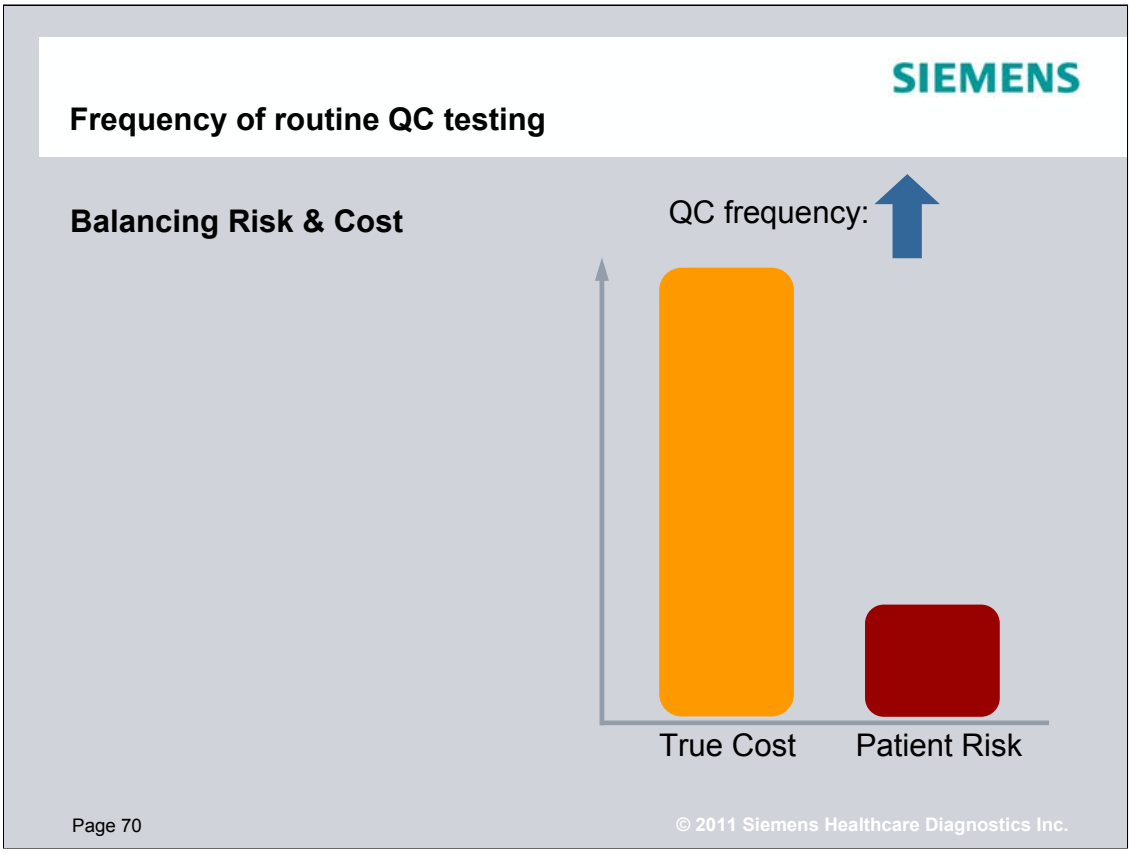
1. Est. frequency of random failure
2. Est. average number of "At Risk" patient results based on frequency of QC
3. Est. Costs:
   - Look back – cost ?
   - Phone calls – how many ? Cost ?
   - Corrected reports – how many ? cost ?
   - Cost of incorrect treatment
   - Liability – analyte dependant

True Cost = Direct cost + Indirect cost

Page 69

Let's look at estimating failure cost. First we need to estimate how often a real failure of the system is likely to occur. This will be fairly infrequent. Remember, the common reasons for changes to system performance are all event based and we are addressing them with our event based QC. Our concern here is the random failure. Then we need to estimate how many patient results are at risk if a failure occurs. Generally the average number of patient results at risk is half the number of results that would likely be reported between any teo routine QC events. Now we look at the costs of following up on those at risk patient results. Based on the lab's protocol, what is the look back process ? How many patient samples are retested, if any ? What is the likelihood of phone calls and corrected reports and estimate the cost. Then we have to factor in some cost for the possibility of incorrect treatment or liability. While an event like this may have huge costs, it will be a rare occurrence, so the cost we factor in can be modest.

Now our true costs is the sum of the direct costs plus the indirect costs and we can play "what if" by looking at varying the frequency of routine QC testing and see what happens to the over all true cost. Lower QC frequency lowers direct and increases indirect. So with a little experimenting using our own testing volumes and protocols we can get an idea how to minimize the true cost.

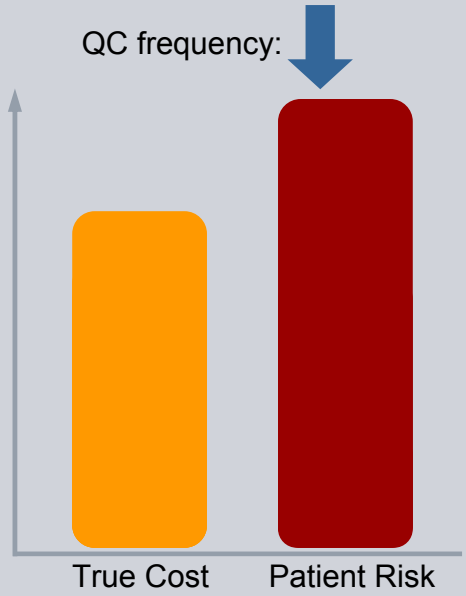


In the end we try to balance the true cost of our QC protocol with patient risk. If we increase the frequency of QC testing, we lower patient risk, but our costs go up.

SIEMENS

Frequency of routine QC testing

Balancing Risk & Cost



Page 71

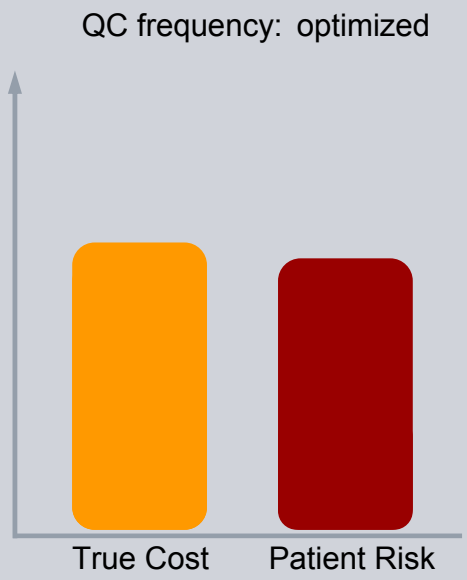
If we decrease QC frequency, we increase patient risk, but our costs go down. However, the costs don't drop as much as we might expect. Decreasing QC frequency lowers direct cost, but increases indirect cost.

SIEMENS

Frequency of routine QC testing

Balancing Risk & Cost

- Optimum will be lab dependant
- No single universal answer
- Note: system reliability not directly a factor



Page 72

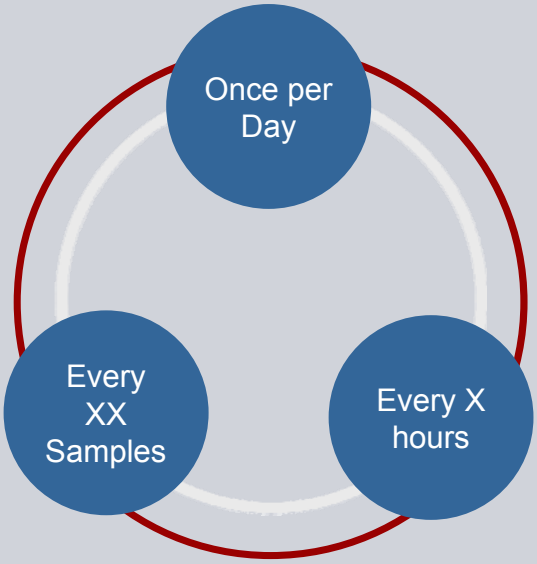
We also recognize that we can never eliminate patient risk no matter how often we test QC samples. So the optimal protocol balances risk and cost and tries to get the most benefit in risk reduction for a true cost that can be sustained.

This optimum will be different for each laboratory. There is no single universally correct answer. We each have to figure it out. Also note that in this discussion, the expected frequency of system failure was not a factor used. That is because once the expected failure rate drops below a threshold, the risk management aspect of the QC protocol becomes more important than the expected frequency of failure.

SIEMENS

How to schedule QC?

1. After all appropriate events – calibration, lot change, etc.
2. Periodically:
   - "Once per day"
     - US regulatory minimum
     - Probably not effective for risk management
   - "Every XX patient samples"
     - Easiest way to estimate risk
     - Practical issues
     - Bracketing
   - "Every X hours"
     - "X" hours set based on risk
     - Probably most practical



Page 73

So we have decided based on true cost and risk management how often we may want to test QC samples. Now how do we implement that ? First, we test QC samples after every event that may alter system performance. Then for the periodic testing, what are the options. The CLIA minimum of once per day is probably not adequate to effectively manage indirect costs and patient risk for most laboratories. Remember, just because we are doing something that is the legal minimum, that doesn't mean we are doing it the best way possible.

Another way we can schedule QC samples is every XX patient samples. This makes it very easy to estimate how many patient samples may be at risk if we have a true failure, but it can be an awkward way to schedule QC. Since testing volume varies widely between analytes, this approach can have us testing QC samples for small groups of different methods quite often. This has a negative impact on workflow and drives up direct costs. This approach is also difficult to use unless QC testing can be auto-scheduled by the instrument, middleware or LIS. Folks working on the instrument cannot possibly keep track of how many samples have been tested for a given method. This approach is the foundation of QC bracketing, which is used in some labs and is mandated for some testing.

Finally there is the way most of us schedule routine QC… every X hours. Using the approach we have discussed we would use our estimates balancing total cost and risk to decide how long a time we should have between each QC event. This is probably the most practical approach because we can keep track of the time interval manually and increasingly instruments, middleware, etc. can auto-schedule QC based on time. If we use the approach we have discussed to determine the optimum time interval, this can be an effective way to do QC.

SIEMENS

Steps to Optimized QC

1. **Decide on the quality goal**
   - What's the Total Allowable Error ?
2. **Evaluate method performance compared to goals**
   - What's the sigma metric ?
3. **Choose the QC rule(s) and frequency**
   - Use Sigma metric to help with rule selection
   - Risk management strategy and method performance guide QC frequency
4. **Set and maintain effective QC targets**
   - The best protocol will not be effective if the targets are not correctly set
5. **Use trends and data to troubleshoot QC rule failures**
   - Well designed QC protocol can help identify problem

Let's review the steps to optimize our QC protocol.

1. We decide on our quality goal – and use total allowable error to help

2. We compare actual method performance to the quality goals and calculate the sigma metric

3. We use the sigma metric to help select the optimal QC rules and the true costs and risk management to determine QC frequency

4. We set and maintain effective QC targets

5 We use the tools built into our QC protocol to help us troubleshoot when we do find a problem

This approach can help us use QC to best advantage assure the highest possible quality while still having a program that is practical and cost effective.