CLINICAL AND LABORATORY STANDARDS INSTITUTE®

4th Edition

C24

Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions

This guideline provides definitions, principles, and approaches to laboratory quality control design, implementation, and assessment.

A guideline for global application developed through the Clinical and Laboratory Standards Institute consensus process.

Clinical and Laboratory Standards Institute

Setting the standard for quality in medical laboratory testing around the world.

The Clinical and Laboratory Standards Institute (CLSI) is a not-for-profit membership organization that brings together the varied perspectives and expertise of the worldwide laboratory community for the advancement of a common cause: to foster excellence in laboratory medicine by developing and implementing medical laboratory standards and guidelines that help laboratories fulfill their responsibilities with efficiency, effectiveness, and global applicability.

Consensus Process

Consensus—the substantial agreement by materially affected, competent, and interested parties—is core to the development of all CLSI documents. It does not always connote unanimous agreement, but does mean that the participants in the development of a consensus document have considered and resolved all relevant objections and accept the resulting agreement.

Commenting on Documents

CLSI documents undergo periodic evaluation and modification to keep pace with advancements in technologies, procedures, methods, and protocols affecting the laboratory or health care.

CLSI's consensus process depends on experts who volunteer to serve as contributing authors and/or as participants in the reviewing and commenting process. At the end of each comment period, the committee that developed the document is obligated to review all comments, respond in writing to all substantive comments, and revise the draft document as appropriate.

Comments on published CLSI documents are equally essential, and may be submitted by anyone, at any time, on any document. All comments are managed according to the consensus process by a committee of experts.

Appeals Process

When it is believed that an objection has not been adequately considered and responded to, the process for appeals, documented in the CLSI Standards Development Policies and Processes, is followed.

All comments and responses submitted on draft and published documents are retained on file at CLSI and are available upon request.

Get Involved—Volunteer!

Do you use CLSI documents in your workplace? Do you see room for improvement? Would you like to get involved in the revision process? Or maybe you see a need to develop a new document for an emerging technology? CLSI wants to hear from you. We are always looking for volunteers. By donating your time and talents to improve the standards that affect your own work, you will play an active role in improving public health across the globe.

For additional information on committee participation or to submit comments, contact CLSI.

Clinical and Laboratory Standards Institute
950 West Valley Road, Suite 2500
Wayne, PA 19087 USA
P: +1.610.688.0100
F: +1.610.688.0700
www.clsi.org
standard@clsi.org

C24, 4th ed.
September 2016
Replaces C24-A3

Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions

Curtis A. Parvin, PhD
Nils B. Person, PhD, FACB
Nikola Baumann, PhD
Lili Duan, PhD
A. Paul Durham
Valerio M. Genta, MD
Jeremie Gras, MD
Greg Miller, PhD
Megan E. Sawchuk, MT(ASCP)

Abstract

Clinical and Laboratory Standards Institute guideline C24—*Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions* discusses the principles of statistical QC, with particular attention to the planning of a QC strategy and the application of statistical QC in a medical laboratory. Although these principles are of interest to manufacturers, this guideline is intended for use by medical laboratory personnel in order to provide a QC strategy that uses control materials that are external to a reagent kit, instrument, or measuring system and that are intended to simulate the measurement of a patient specimen.

Clinical and Laboratory Standards Institute (CLSI). *Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions.* 4th ed. CLSI guideline C24 (ISBN 1-56238-946-7 [Print]; ISBN 1-56238-947-5 [Electronic]). Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087 USA, 2016.

Suggested Citation

CLSI. *Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions*. 4th ed. CLSI guideline C24. Wayne, PA: Clinical and Laboratory Standards Institute; 2016.

Previous Editions:
March 1985, September 1986, May 1991, February 1999, June 2006

ISBN 1-56238-946-7 (Print)
ISBN 1-56238-947-5 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic) Volume 36, Number 12

Committee Membership

Consensus Council

Carl D. Mottram, RRT, RPFT, FAARC
Chairholder
Mayo Clinic
USA

J. Rex Astles, PhD, FACB, DABCC
Centers for Disease Control and Prevention
USA

Lucia M. Berte, MA, MT(ASCP)SBB, DLM; CQA(ASQ)CMQ/OE
Laboratories Made Better!
USA

Karen W. Dyer, MT(ASCP), DLM
Centers for Medicare & Medicaid Services
USA

Dennis J. Ernst, MT(ASCP), NCPT(NCCT)
Center for Phlebotomy Education
USA

Thomas R. Fritsche, MD, PhD, FCAP, FIDSA
Marshfield Clinic
USA

Mary Lou Gantzer, PhD, FACB
BioCore Diagnostics
USA

Loralie J. Langman, PhD
Mayo Clinic
USA

Joseph Passarelli
Roche Diagnostics Corporation
USA

James F. Pierson-Perry
Siemens Healthcare Diagnostics Inc.
USA

Andrew Quintenz
Bio-Rad Laboratories, Inc.
USA

Robert Rej, PhD
New York State Department of Health – Wadsworth Center
USA

Zivana Tezak, PhD
FDA Center for Devices and Radiological Health
USA

Document Development Committee on Statistical QC for Clinical Chemistry

Curtis A. Parvin, PhD
Chairholder
Bio-Rad Laboratories, Inc.
USA

Nils B. Person, PhD, FACB
Vice-Chairholder
Siemens Healthcare Diagnostics, Inc.
USA

Nikola Baumann, PhD
Mayo Clinic
USA

Lili Duan, PhD
FDA Center for Devices and Radiological Health
USA

Jeremie Gras, MD
Clinique St. Luc
Belgium

Greg Miller, PhD
Virginia Commonwealth University Health System
USA

Megan E. Sawchuk, MT(ASCP)
Centers for Disease Control and Prevention
USA

Staff

Clinical and Laboratory Standards Institute
USA

Luann Ochs, MS
Project Manager

Megan L. Tertel, MA, ELS
Editorial Manager

Joanne P. Christopher, MA, ELS
Editor

Laura Martin
Editor

Michael A. Russell, MA
Editor

Acknowledgment for the Expert Panel on Clinical Chemistry and Toxicology

CLSI, the Consensus Council, and the Document Development Committee on Statistical QC for Clinical Chemistry gratefully acknowledge the Expert Panel on Clinical Chemistry and Toxicology for serving as technical advisors and subject matter experts during the development of this guideline.

Expert Panel on Clinical Chemistry and Toxicology

Johanna Camara, PhD
Chairholder
National Institute of Standards and Technology
USA

Lorin M. Bachmann, PhD, DABCC
Vice-Chairholder
Virginia Commonwealth University Health System
USA

Karl De Vore
Bio-Rad Laboratories, Inc.
USA

Lili Duan, PhD
FDA Center for Devices and Radiological Health
USA

Kamisha Johnson-Davis, PhD, DABCC, FACB
University of Utah and ARUP Laboratories
USA

Gregory T. Maine, PhD, FACB
Abbott
USA

Godwin Ogbonna, PhD
Ortho-Clinical Diagnostics, Inc.
USA

Curtis Oleschuk, PhD, FCACB
Diagnostic Services of Manitoba
Canada

David B. Sacks, MB, ChB, FRCPath
National Institutes of Health
USA

Acknowledgment

CLSI, the Consensus Council, and the Document Development Committee on Statistical QC for Clinical Chemistry gratefully acknowledge the following volunteers for their important contributions to the development of this guideline:

A. Paul Durham
APD Consulting
USA

Valerio M. Genta, MD
Sentara Virginia Beach General Hospital
USA

Contents

Contents (Continued)

Foreword

The medical laboratory community has used C24, now in its fourth edition, for more than 20 years. Today, statistical QC is still critically important to ensure the quality of the results of any laboratory measurement procedure. The almost universal applicability of statistical QC to quantitative measurement procedures provides laboratories with an essential quality management tool that can be used to monitor the effects of many instrument, reagent, environment, and operator variables on the outcome of a measurement process.

The laboratory director is generally responsible for the laboratory QC program. The definition of quality requirements for the tests being performed is particularly important because laboratory managers, supervisors, scientists, and quality specialists often use those quality requirements to select and validate appropriate measurement and control procedures. C24's approach provides medical laboratory scientists with practical guidance on how to satisfy recommendations by authorities and/or accreditation organizations.[1]

The concepts, approaches, and practices discussed in this guideline are interdependent and all should be carefully studied and considered when developing the specific QC strategy for any measurement procedure, system, or laboratory. C24 highlights the technical issues that need a careful scientific approach to designing, implementing, and assessing QC strategies in order for laboratories to achieve the quality requirements needed by the physicians and patients they serve.

Overview of Changes

This guideline replaces the previous edition of the approved guideline, C24-A3, published in 2006. The fourth edition maintains the focus on principles and approaches to laboratory QC design, implementation, and assessment that reflect the realities of the modern medical laboratory and its role within the health care enterprise. Several changes were made in this edition, including:

- The alignment of principles and definitions to be consistent with and to supplement the general patient risk model described in CLSI document EP23™[2]

- The introduction of additional performance measures useful for evaluating the performance characteristics of a QC strategy (see Chapter 5)

- Expanded guidance on setting target values and SDs for QC materials (see Subchapter 5.3)

- A greater focus on QC frequency and QC schedules as a critical part of a QC strategy (see Subchapter 5.5)

- A substantive chapter on recovering from an out-of-control condition (see Chapter 6), including sections on:
  - Responding to an out-of-control QC event
  - Responding to an out-of-control condition
  - Identifying and correcting reported erroneous patient results

NOTE: The content of this guideline is supported by the CLSI consensus process, and does not necessarily reflect the views of any single individual or organization.

Key Words

Patient risk, quality control, quality control plan, quality control rules, quality control strategy, quality requirements, Sigma metric

Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions

Chapter 1: Introduction

This chapter includes:

- Guideline's scope and applicable exclusions
- Background information pertinent to the guideline's content
- Standard precautions information
- "Note on Terminology" that highlights particular use and/or variation in use of terms and/or definitions
- Terms and definitions used in the guideline
- Abbreviations and acronyms used in the guideline

1.1 Scope

This guideline explains the purpose of statistical QC for quantitative measurement procedures, describes an approach for planning a QC strategy for a particular measurement procedure, describes the use of QC material and QC data, and provides examples that demonstrate a practical QC planning process for medical laboratories.

The recommendations for establishing and maintaining a statistical QC strategy are applicable to quantitative laboratory measurement procedures in all fields of laboratory medicine for which stable control materials can be measured in the same manner as patient specimens. The intended users of this guideline include those responsible for designing, implementing, and using QC, ie, medical laboratory scientists.

This guideline does **not:**

- Describe built-in control mechanisms that might be part of a measuring system, or qualitative or semiquantitative measurement procedures.
- Define specific QC strategies that are appropriate for an individual device or technology.
- Describe alternatives to statistical process control, eg, real-time patient-based QC.
- Consider specific legal requirements that may impose different philosophies or procedures on QC practices (eg, a specific approach for defining quality requirements, specific values for quality requirements, a specific procedure for determining target values for control materials, or a frequency and number of QC measurements) defined by government regulation in a specific country or region.

Additionally, there are types of random errors that may affect measurements performed on individual specimens, rather than a whole group of specimens, and those errors are not detected by a statistical QC

strategy. Such errors may be due to the specific design of a measuring system (eg, effect of specimen viscosity, carryover from a previous specimen, or specimen-specific interferences) or possible operator errors that affect individual specimens, as well as preexamination errors of specimen preparation, storage, and transportation. Special QC strategies may be needed to monitor known special vulnerabilities that relate to a particular device or system design.

1.2 Background

Statistical QC strategies are implemented to monitor a measurement procedure's performance to detect any change relative to stable baseline analytical performance. When the actual performance deviates from the expected model, the QC strategy is designed to alert the laboratorian to a change that may affect medical decision making and potentially lead to incorrect treatment, delays in treatment, or patient harm. Designing an effective QC strategy entails determining the magnitude of the change in performance that compromises the usefulness of the measurement procedure results.

There is abundant literature explaining the theoretical and practical bases for initiating and maintaining QC strategies in clinical chemistry[3-9]; however, the routine practice of statistical QC depends on understanding how to:

- Plan QC strategies based on the performance of the measurement procedure and the performance needed to support the intended medical use of the results, including selecting appropriate control materials, establishing the expected values for those control materials, determining when to evaluate controls, and identifying the control rules to determine acceptable performance.

- Implement QC strategies to identify situations when a measurement procedure may not be providing results that are suitable for use in medical decisions.

- Respond to out-of-control situations.

The prevalence of a broad range of automated medical laboratory instruments using widely different measuring principles has complicated the terminology and the steps necessary for establishing QC strategies. There are some highly automated systems that can perform specific, built-in checks that help detect potential problems and alert the operator to instrument malfunction. However, the benefit of statistical QC using samples intended to simulate measurement of patient specimens is that it monitors the outcome of many of the variables and steps that occur in the entire measurement procedure.

1.3 Standard Precautions

Because it is often impossible to know what isolates or specimens might be infectious, all patient and laboratory specimens are treated as infectious and handled according to "standard precautions." Standard precautions are guidelines that combine the major features of "universal precautions and body substance isolation" practices. Standard precautions cover the transmission of all known infectious agents and thus are more comprehensive than universal precautions, which are intended to apply only to transmission of bloodborne pathogens. Published guidelines are available that discuss the daily operations of diagnostic medicine in humans and animals while encouraging a culture of safety in the laboratory.[10] For specific precautions for preventing the laboratory transmission of all known infectious agents from laboratory instruments and materials and for recommendations for the management of exposure to all known infectious diseases, refer to CLSI document M29.[11]

1.4 Terminology

1.4.1 A Note on Terminology

CLSI, as a global leader in standardization, is firmly committed to achieving global harmonization whenever possible. Harmonization is a process of recognizing, understanding, and explaining differences while taking steps to achieve worldwide uniformity. CLSI recognizes that medical conventions in the global metrological community have evolved differently in different countries and regions, and that legally required use of terms, regional usage, and different consensus timelines are all important considerations in the harmonization process. CLSI recognizes its important role in these efforts, and its consensus process focuses on harmonization of terms to facilitate the global application of standards and guidelines.

1.4.2 Definitions

accuracy (of measurement) – closeness of agreement between a measured quantity value and a true quantity value of a measurand[12]; **NOTE 1:** The concept "measurement accuracy" is not a quantity and is not given a numerical quantity value. A measurement is said to be more accurate when it offers a smaller measurement error[12]; **NOTE 2:** The term "measurement accuracy" should not be used for measurement trueness and the term "measurement precision" should not be used for "measurement accuracy", which, however, is related to both these concepts[12]; **NOTE 3:** "Measurement accuracy" is sometimes understood as closeness of agreement between measured quantity values that are being attributed to the measurand.[12]

allowable total error (TEa) – an analytical quality goal that sets a limit for both the imprecision (random error) and bias (systematic error) that are tolerable in a single measurement or single test result; **NOTE 1:** For quality control (QC) planning, it is assumed there are no specimen-specific influences because they are a component of overall method performance that is not monitored by a statistical QC strategy; **NOTE 2:** Some publications denote allowable total error as "ATE."

analyte – constituent of a sample with a measurable property[13]; **NOTE:** In "mass of protein in 24-hour urine," "protein" is the analyte and "mass" is the property. In "concentration of glucose in plasma," "glucose" is the analyte and "concentration" is the property. In both cases, the full phrase represents the measurand.[13]

bias (of measurement) – estimate of a systematic measurement error[12]; difference between the expectation of a test result or measurement result and a true value[14]; **NOTE 1:** In practice, the accepted reference value is substituted for the true value[14]; **NOTE 2:** Bias represents the quantitative expression of trueness.

coefficient of variation (CV) – (positive random variable) standard deviation (SD) divided by the mean[15]; **NOTE 1:** The CV is commonly reported as a percentage[15]; **NOTE 2:** The predecessor term "relative SD" is deprecated by the term CV.[15]

control limit – the most extreme value of a quality control material that is still considered to be acceptable.

erroneous result – a patient result that fails its quality requirement; **NOTE 1:** The quality requirement is usually expressed in terms of an allowable total error (TEa) requirement. If the measurement error in a patient's result exceeds the TEa requirement, the result is erroneous; **NOTE 2:** May also be referred to as an incorrect result or an unacceptable result.

error (of measurement) – measured quantity value minus a reference quantity value[12]; **NOTE 1:** The concept of "measurement error" can be used both a) when there is a single reference quantity value to

refer to, which occurs if a calibration is made by means of a measurement standard with a measured quantity value having a negligible measurement uncertainty or if a conventional quantity value is given, in which case the measurement error is known, and b) if a measurand is supposed to be represented by a unique true quantity value or a set of true quantity values of negligible range, in which case the measurement error is not known[12]; **NOTE 2:** Measurement error should not be confused with production error or mistake.[12]

imprecision – the random dispersion of a set of replicate measurements and/or values expressed quantitatively by a statistic; **NOTE:** It is expressed numerically as standard deviation or **coefficient of variation.**

mean (arithmetic)//average – sum of random variables in a random sample divided by the number of terms in the sum[15]; **NOTE:** The sample mean considered as a statistic is often used as an estimator for the population mean. A common synonym is arithmetic mean.[15]

measurand – quantity intended to be measured[12]; **NOTE 1:** The specification of a measurand requires knowledge of the kind of quantity, description of the state of the phenomenon, body, or substance carrying the quantity, including any relevant component, and the chemical entities involved; **NOTE 2:** In the second edition of the VIM[12] and in IEC 60050-300:2001,[16] the measurand is defined as the "particular quantity subject to measurement"[12]; **NOTE 3:** The measurement, including the measuring system and the conditions under which the measurement is carried out, might change the phenomenon, body, or substance such that the quantity being measured may differ from the measurand as defined. In this case, adequate correction is necessary; **NOTE 4:** In chemistry, "analyte," or the name of a substance or compound, are terms sometimes used for "measurand." This usage is erroneous because these terms do not refer to quantities.[12]

measurement procedure – detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result[12]; **NOTE 1:** A measurement procedure is usually documented in sufficient detail to enable an operator to perform a measurement[12]; **NOTE 2:** A measurement procedure can include a statement concerning a target measurement uncertainty[12]; **NOTE 3:** Formerly, the term "analytical method" was used in C24.

measuring interval – set of values of quantities of the same kind that can be measured by a given measuring instrument or measuring system with specified instrumental measurement uncertainty, under defined conditions[12]; **NOTE 1:** In some fields, the term is "analytical measurement range," "measuring range," or "measurement range"; **NOTE 2:** The lower limit of a measuring interval should not be confused with detection limit.[12]

out-of-control condition – a process or component of a process that is not operating in its stable state; **NOTE 1:** For quantitative measurement procedures, an out-of-control condition is usually described in terms of a shift or drift away from the stable mean of the measurement procedure, or as an increase in random imprecision above the stable imprecision of the measurement procedure; **NOTE 2:** May be referred to as an out-of-control error condition, or error condition.

precision (of measurement) – closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions[12]; **NOTE 1:** Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement[12]; **NOTE 2:** The "specified conditions" can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement[12]; **NOTE 3:** Measurement precision is used to define measurement repeatability, intermediate measurement precision,

and measurement reproducibility[12]; **NOTE 4:** Sometimes "measurement precision" is erroneously used to mean "measurement accuracy."[12]

proficiency testing (PT)//external quality assessment (EQA) – a program in which multiple samples are periodically sent to members of a group of laboratories for analysis and/or identification, in which each laboratory's results are compared with those of other laboratories in the group and/or with an assigned value, and reported to the participating laboratory and others; **NOTE 1:** Used to establish between-laboratory and between-instrument comparability that is, if possible, in agreement with a reference standard (when one exists). EQA schemes may be regional, national, or international. EQA is sometimes also referred to as PT, especially when the external agency is a regulatory agency; **NOTE 2:** Interlaboratory comparisons and other performance evaluations that may extend throughout all phases of the testing cycle, including interpretation of results; determination of individual and collective laboratory performance characteristics of examination procedures by means of interlaboratory comparison; **NOTE 3:** The primary objectives of PT/EQA are educational and may be supported by additional elements.

quality control (QC) – part of quality management focused on fulfilling quality requirements[17]; **NOTE 1:** In health care testing, the set of procedures based on measurement of a stable material that is similar to the intended patient specimen, to monitor the ongoing performance of a measurement procedure and detect change in that performance relative to stable baseline analytical performance; **NOTE 2:** QC includes testing QC materials, charting the results and analyzing them to identify sources of error, and evaluating and documenting any remedial action taken as a result of this analysis.

quality control (QC) event – the occurrence of one or more QC measurements and a QC rule evaluation using the QC results; **NOTE:** This may also be referred to as a QC evaluation.

quality control (QC) plan – a document that describes the practices, resources, and sequences of specified activities to control the quality of a particular measuring system or measurement procedure to ensure requirements for its intended purpose are met.

quality control (QC) result – the quantity obtained from a QC measurement.

quality control (QC) rule – decision criteria used in the process of deciding whether a measurement procedure is operating within its stable (in-control) state.

quality control (QC) rule evaluation – the process of deciding whether a measurement procedure is operating in its stable (in-control) state by applying a QC rule to a set of QC results.

quality control (QC) strategy – the number of QC materials to measure, the number of QC results and the QC rule to use at each QC event, and the frequency of QC events; **NOTE:** May also be referred to as QC procedure.

quality requirement – specification of the characteristics necessary for a product or service to be fit for its intended use; **NOTE:** For a laboratory measurement procedure, the quality requirement is usually expressed in terms of an allowable total error (TEa). If the measurement error in a patient's result exceeds the TEa, the result fails to meet its quality requirement.

reference quantity value//reference value – quantity value used as a basis for comparison with values of quantities of the same kind[12]; **NOTE 1:** A reference quantity value can be a true quantity value of a measurand, in which case it is unknown, or a conventional quantity value, in which case it is known[12]; **NOTE 2:** A reference quantity value with associated measurement uncertainty is usually provided with reference to a) a material, eg, a certified reference material, b) a device, eg, a stabilized laser, c) a reference measurement procedure, or d) a comparison of measurement standards[12]; **NOTE 3:** An "accepted reference value" is a value that serves as an agreed-upon reference for comparison, and which

is derived as a theoretical or established value, based on scientific principles; an assigned or certified value, based on experimental work of some national or international organization; or a consensus or certified value, based on collaborative experimental work under the auspices of a scientific or engineering group.[18]

sample – collection of one or more parts initially taken from a system and intended to provide information about the system, or to serve as a basis for a decision about the system[19]; **NOTE 1:** A sample is prepared from the patient specimen and used to obtain information by means of a specific laboratory test; **NOTE 2:** The system from which a sample is taken may not be of the same type as that of the measurand. For example, a given blood sample may serve for measurement of the pH (negative logarithm of hydrogen ion concentration) in plasma, or for measurement of the hemoglobin concentration in erythrocytes; **NOTE 3:** For the purposes of this guideline, the term "sample" is used to denote nonhuman or modified human materials such as quality control materials, calibrators, or proficiency testing/external quality assessment materials.

specimen – (patient) discrete portion of a body fluid or tissue taken for examination, study, or analysis of one or more quantities or characteristics to determine the character of the whole.[13]

stability (of a measuring instrument) – property of a measuring instrument, whereby its metrological properties remain constant in time[12]; **NOTE:** Stability may be quantified in several ways[12]; **EXAMPLE 1:** In terms of the duration of a time interval over which a metrological property changes by a stated amount[12]; **EXAMPLE 2:** In terms of the change of a property over a stated time interval.[12]

stable process – process in a state of statistical control[14]; **NOTE 1:** A stable process will generally behave as though the samples from the process at any time are simple random samples from the same population[14]; **NOTE 2:** This state does not imply that the random variation is large or small, within or outside of specification, but rather that the variation is predictable using statistical techniques[14]; **NOTE 3:** The process capability of a stable process is usually improved by fundamental changes that reduce or remove some of the random causes present and/or adjusting the mean towards the preferred value[14]; **NOTE 4:** In some processes, the mean of a characteristic can have a drift or the standard deviation (SD) can increase due, for example, to wear-out of tools or depletion of concentration in a solution. A progressive change in the mean or SD of such a process is considered due to systematic and not random causes. The results, then, are not simple random samples from the same population.[14]

statistical process control – activities focused on the use of statistical techniques to reduce variation, increase knowledge about the process, and steer the process in the desired way.[14]

systematic measurement error//systematic error (of measurement) – component of measurement error that in replicate measurements remains constant or varies in a predictable manner[12]; **NOTE 1:** Systematic measurement error, and its causes, can be known or unknown[12]; **NOTE 2:** Systematic measurement error equals measurement error minus random measurement error.[12]

true quantity value//true value – quantity value consistent with the definition of a quantity[12]; **NOTE 1:** There are multiple approaches to considering the true value; **NOTE 2:** In the Error Approach to describing measurement, a true quantity value is considered unique and, in practice, unknowable. The Uncertainty Approach is to recognize that, owing to the inherently incomplete amount of detail in the definition of a quantity, there is not a single true quantity value but rather a set of true quantity values consistent with the definition. However, this set of values is, in principle and in practice, unknowable. Other approaches dispense altogether with the concept of true quantity value and rely on the concept of metrological compatibility of measurement results for assessing their validity[12]; **NOTE 3:** In the special case of a fundamental constant, the quantity is considered to have a single true quantity value[12]; **NOTE 4:** When the definitional uncertainty associated with the measurand is considered to be negligible compared to the other components of the measurement uncertainty, the measurand may be considered to have an

"essentially unique" true quantity value. This is the approach taken by the GUM[20] and its associated documents, where the word "true" is considered to be redundant.[12]

uncertainty (of measurement) – non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used[12]; **NOTE 1:** Measurement uncertainty includes components arising from systematic effects, such as components associated with corrections and the assigned quantity values of measurement standards, as well as the definitional uncertainty. Sometimes estimated systematic effects are not corrected for but, instead, associated measurement uncertainty components are incorporated[12]; **NOTE 2:** The parameter may be, for example, a standard deviation (SD) called standard measurement uncertainty (or a specified multiple of it), or the half-width of an interval, having a stated coverage probability[12]; **NOTE 3:** Measurement uncertainty comprises, in general, many components. Some of these may be evaluated by Type A evaluation of measurement uncertainty from the statistical distribution of the quantity values from series of measurements and can be characterized by SDs. The other components, which may be evaluated by Type B evaluation of measurement uncertainty, can also be characterized by SDs, evaluated from probability density functions based on experience or other information[12]; **NOTE 4:** In general, for a given set of information, it is understood that the measurement uncertainty is associated with a stated quantity value attributed to the measurand. A modification of this value results in a modification of the associated uncertainty.[12]

validation – confirmation, through the provision of objective evidence, that requirements for a specific intended use or application have been fulfilled.[17]

verification – confirmation, through the provision of objective evidence, that specified requirements have been fulfilled.[17]

1.4.3 Abbreviations and Acronyms

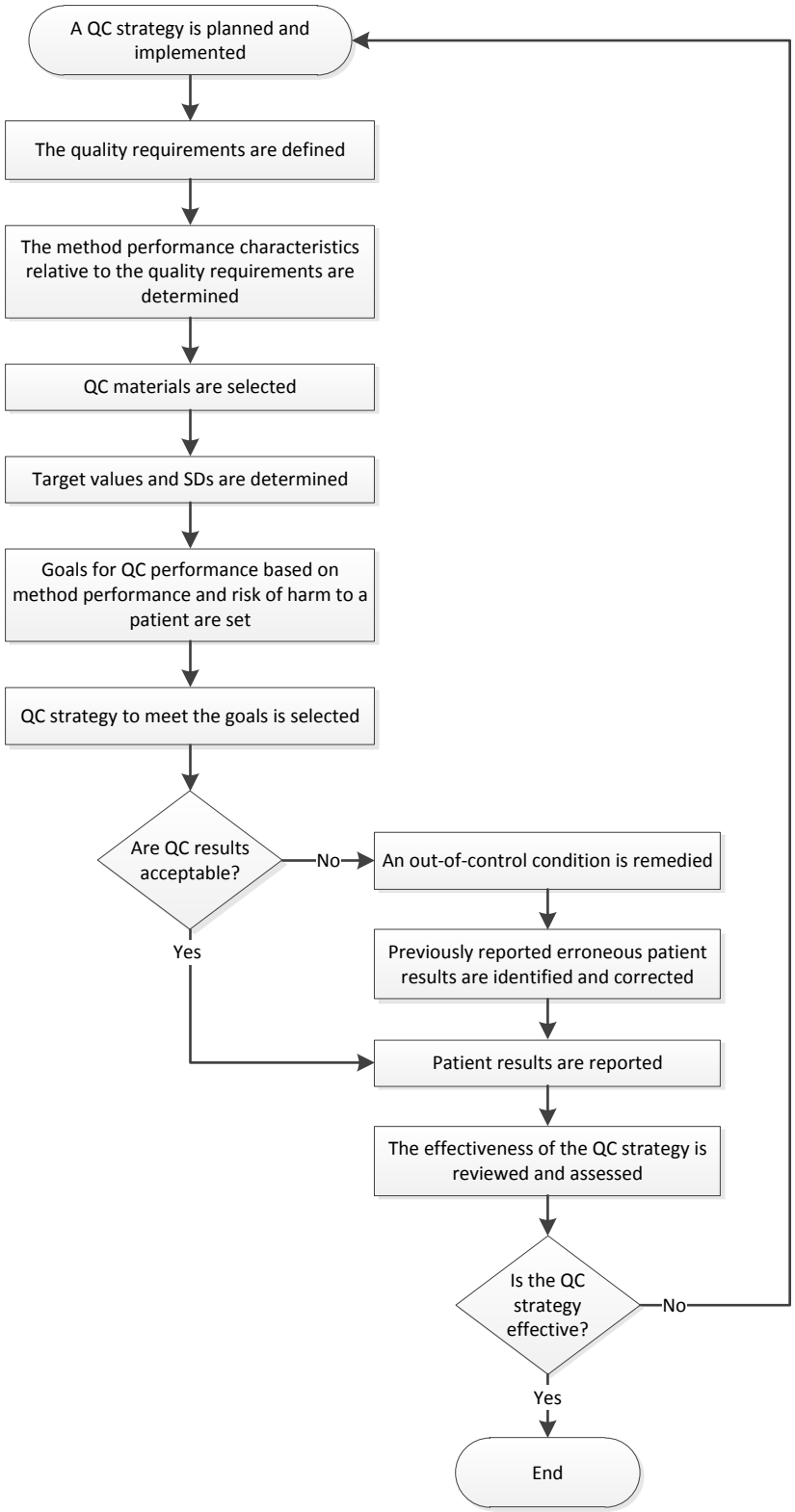
CUSUM	cumulative sum
CV	coefficient of variation
EQA	external quality assessment
EWMA	exponentially weighted moving average
PT	proficiency testing
QC	quality control
SD	standard deviation
SDI	standard deviation interval
TEa	allowable total error
TSH	thyroid-stimulating hormone

This page is intentionally left blank.

Chapter 2: Path of Workflow

This chapter includes:

- The process flow chart for planning and implementing a QC strategy (see Figure 1)



* Five basic symbols are used in process flow charts: Oval (signifies the beginning or end of a process), Arrow (connects process activities), Box (designates process activities), Diamond (includes a question with alternative "Yes" and "No" responses), Pentagon (signifies another process).
Abbreviations: QC, quality control; SD, standard deviation.

Figure 1. Process Flow Chart for Planning and Implementing a QC Strategy*

Chapter 3: Purpose of Statistical Quality Control

This chapter includes:

- Understanding how QC relates to patient risk
- Understanding the quality requirements for a measurement procedure
- Determining whether a measurement procedure is meeting its quality requirements
- Understanding the types of out-of-control conditions

The purpose of statistical QC in the medical laboratory, as part of the statistical control process, is to identify as quickly as possible any change in the stable operation of a measurement procedure that causes a significant increase in the risk of producing and reporting erroneous patient results that could adversely affect medical decision making. Some important points to consider are:

- Statistical QC monitors a laboratory measurement procedure, but it should be planned with the patient in mind. What constitutes an important change in a measurement procedure and how quickly such a change needs to be detected should be based on the patient risk implications.

- Patient risk depends on the likelihood that inappropriate medical decisions or actions may occur based on erroneous laboratory results. In order to assess the patient risk implications of a change in the stable operation of the measurement procedure, it is necessary to define the total amount of error in a result that is likely to lead to inappropriate decisions.

- Statistical QC is designed to detect a meaningful change in the measurement procedure irrespective of the particular failure mode that caused the change. Failure modes in the measurement procedure that could affect the measurement error in a patient's result are expected to affect the QC material in a similar way.

- Statistical QC testing can also be used to identify opportunities for improvement of the measurement process.

3.1 Quality Control and Patient Risk

The key goal of any laboratory QC plan is to reduce the risk of harm to a patient due to an erroneous result. Although statistical QC is the principle focus of this guideline, it needs to be viewed as one part of an overall quality management plan. The use of a risk management approach to develop a laboratory QC plan is described elsewhere (see CLSI document EP23[2]). When using a risk management approach to develop a QC strategy, three aspects of the failure causing erroneous patient results should be considered:

- How likely it is for the failure to occur (probability)
- How severe the potential harm to the patient is if the failure goes undetected (severity)
- How reliably the QC strategy can detect the failure if it occurs (detectability)

The laboratory's role in causing patient harm relates to reporting of erroneous patient results not fit for their intended use. Laboratory QC is designed to limit the number of erroneous patient results the laboratory reports because of the occurrence of an out-of-control measurement condition. Depending on the measurand and the patient population, the likelihood an erroneous result leads to an inappropriate decision or action that causes patient harm, as well as the severity of that harm, can vary. The laboratory's tolerance for reporting erroneous results should depend on an assessment of the risk of harm. The higher the likelihood that an erroneous result will cause patient harm or the more severe the patient harm, the

more stringent the laboratory should be when identifying an out-of-control condition in order to minimize the number of erroneous results reported.

The risk to patient safety increases when the QC strategy does not detect an out-of-control condition that has medical consequences. For an out-of-control condition to cause harm, an erroneous patient result is reported and an inappropriate medical decision (action or inaction) is made. Examples of situations that can cause harm are:

- The QC strategy did not detect the out-of-control condition.
- The QC strategy detected the out-of-control condition sometime after it affected patient results.
- The response to a QC false rejection caused a delay in reporting results that affected decisions regarding patient management.

A well-designed QC strategy should reliably detect changes in measurement procedure performance that may cause a risk of harm to a patient based on the intended medical use of the results, and it should detect those changes quickly enough to minimize the number of patient results affected (see Subchapter 4.2.3). The goal is to use a QC strategy that can detect change in performance reliably before the clinical quality requirement is exceeded while also minimizing the frequency of false rejections. Minimizing the number of potentially affected patient results is achieved by an appropriate frequency for measuring and evaluating QC samples. Chapter 5 discusses planning a QC strategy in more detail.

3.2 Quality Requirements

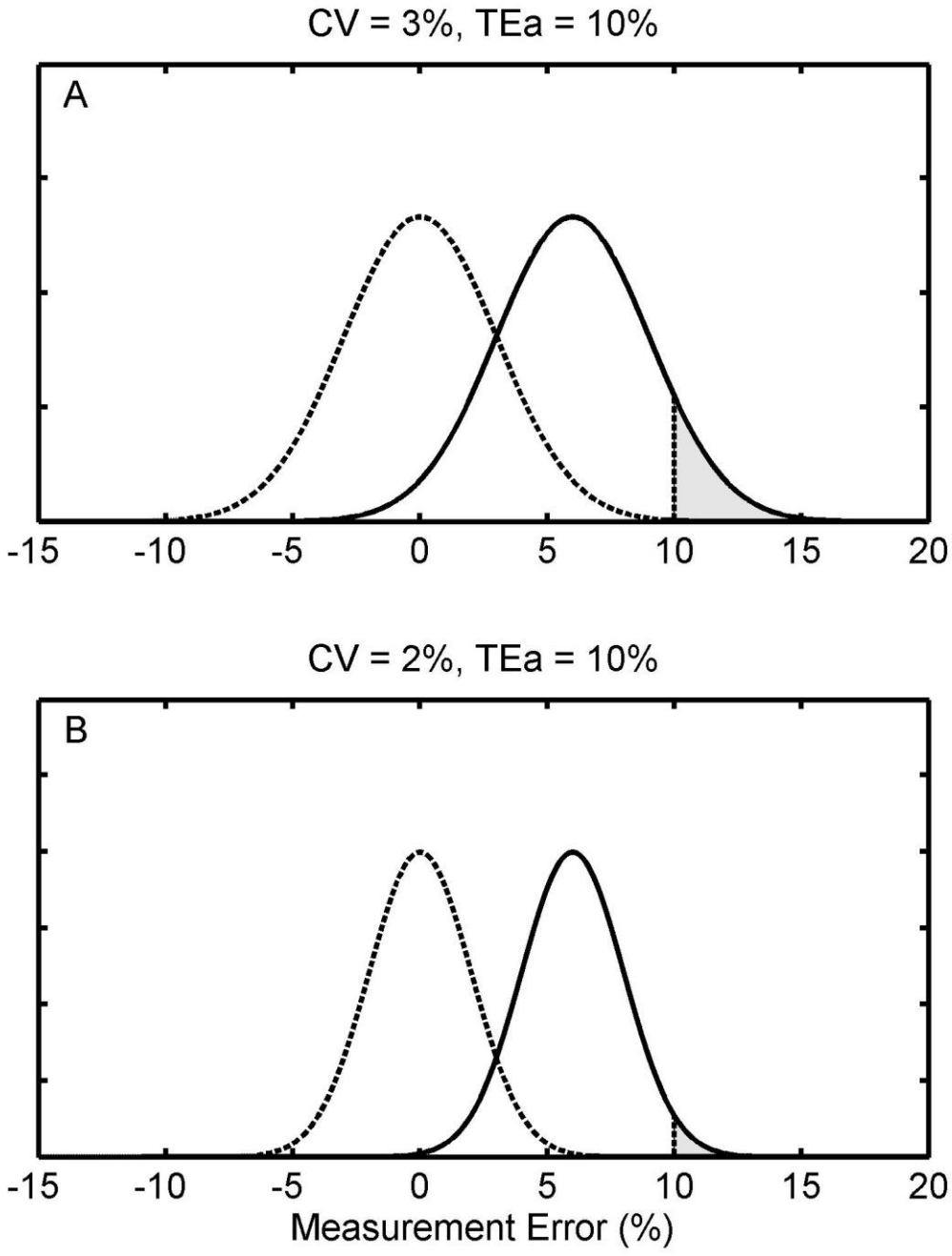
Measurement procedures should be selected that have performance specifications adequate to meet the intended medical use of the results. The allowable total error (TEa), is a commonly used parameter to establish the medical quality requirement.[3] TEa establishes the maximum error that is tolerated without affecting medical decision making and so establishes the "error budget" for a given measurement procedure. There are no universally accepted criteria for defining the magnitude of error that influences clinical decisions. Therefore, the laboratory director should determine TEa limits based on how measurement procedure results are used medically in the population served by the laboratory.

The TEa should be established for each measurand. Because TEa is determined by the medical use of the results for a measurand, it is established independently of the measurement procedure's actual performance characteristics. Additionally, the TEa may be different for the same measurement procedure at different locations because of varying patient needs. Laboratory directors depend heavily on the availability of published information for each measurand to determine the TEa. Sources include clinical studies, biological variability data, and professional practice guidelines or recommendations. A more detailed discussion of establishing TEa is in Subchapter 5.1.

During stable analytical performance, the likelihood that measurement errors exceed TEa should be low to ensure the performance of the measurement procedure can be effectively monitored using statistical QC techniques. If the distribution of measurement variability during stable operation is barely within the TEa limits, then a small change in measurement procedure performance (which is difficult to detect using statistical QC techniques) can cause the likelihood that measurement errors exceed TEa to become unacceptably high. Conversely, if the measurement variability during stable operation is small compared to TEa, then large changes in measurement procedure performance (easy to detect with QC) would be needed for the likelihood of measurement errors exceeding TEa to become unacceptably high.

For example, if TEa is 10% and the stable analytical imprecision of the measurement procedure has a CV of 3%, then during stable operation, the likelihood of measurement errors exceeding TEa is about one in every 1165 measurements. If a 6% shift occurred (a shift equal to twice the stable analytical

imprecision), the likelihood of measurement errors exceeding TEa increases to about 9% of measurements (see Figure 2A). On the other hand, if the stable analytical imprecision of the measurement procedure is CV = 2%, then during stable operation, the likelihood of measurement errors exceeding TEa is less than 1 in every 1.7 million measurements, and an out-of-control shift of 6% (a shift equal to three times the stable analytical imprecision) increases the likelihood of measurement errors exceeding TEa to only about 2% of measurements (see Figure 2B).



Abbreviations: CV, coefficient of variation; TEa, allowable total error.

Figures 2A and 2B. Illustration of the Effect of Shift in Measurement Error Under Different CV Conditions at the Same TEa. The shaded areas represent the fraction with error that exceeds TEa.

C24, 4th ed.

The dashed curves in Figures 2A and 2B represent the distributions of measurement errors (which would be reflected in QC results) expected for a stable operating condition, and the solid curves represent the distributions of measurement errors after a 6% shift in the measurement procedure.

One approach to characterizing the stable performance of a measurement procedure (imprecision and bias) relative to the measurement error quality requirement (TEa) involves the calculation of an index commonly called the Sigma metric. Use of this index is discussed in Subchapter 3.3.3.

3.3 Method Performance Relative to Quality Requirements

Measurement procedure error in the context of statistical QC has typically been considered as made up of two components: constant error, or bias; and random error, or imprecision. Comparison of the expected error associated with stable analytical performance to the clinically based goals can be done separately for each component or for the combination.

3.3.1 Bias

Bias is an estimate of systematic measurement error. Assessing bias relative to performance goals can be challenging. There are three ways to assess bias with regard to developing QC strategies.

The optimal method is to compare results obtained from fresh patient specimens using the measurement procedure and a reference measurement procedure (see CLSI document EP09[21]). Unfortunately, this approach is impractical for most laboratories. Reference measurement procedures have not been developed for many measurands reported in medical laboratories. In addition, reference measurement procedures can be difficult to set up and maintain and are generally not practical for routine testing. In some cases, laboratories have set up reference measurement procedures and accept specimens or share specimens to allow bias estimation. However, recognized reference laboratories[22] are not common and do not support all measurands. Bias can also be assessed with a recovery experiment, such as the approach described in CLSI document EP15.[48] However, reference materials can also be difficult to obtain. Consequently, estimating actual or true bias is difficult and often impossible.

A second approach is to assess relative bias. Often laboratories perform comparison studies when implementing a new measurement procedure. These studies provide information on the relative difference between the new measurement procedure and the one being replaced. Although useful information, these estimates generally do not provide a good estimate of the actual bias of the new measurement procedure primarily because the bias of the comparative measurement procedure is usually unknown. Another way to assess relative bias is comparing the laboratory's results to a peer group mean based on interlaboratory QC data or proficiency testing (PT)/external quality assessment (EQA).[23] Users of interlaboratory QC data should be aware of the possible limitations of interlaboratory QC programs, including statistical methods used to generate the data and the number of laboratories participating. The most commonly used QC data and PT/EQA results frequently show matrix-related bias compared to testing fresh patient specimens, thus obscuring the true patient specimen bias. Consequently, only the laboratory's apparent difference relative to other laboratories using the same measurement procedure is evaluated and does not account for any inherent bias in the peer group measurement procedure. For some measurement procedures, this matrix-related bias can change with new reagent lots so the apparent bias may change from one reagent lot to the next. In this case, peer data composed of results from multiple reagent lots may not be truly representative. When available, PT/EQA programs using samples demonstrated to be commutable with fresh patient specimens may give reasonable assessments of measurement procedure bias.[23] Although programs using commutable samples are not readily available for all measurands, they are becoming more widely available.

Laboratories may have more than one of the same measurement procedure performing patient examinations. In this situation, an individual measurement procedure's relative bias may be defined in

terms of its bias to the group mean of the multiple measurement procedures in the laboratory. As with relative bias compared to a peer group of laboratories, the relative bias of an individual measurement procedure compared to the laboratory's group mean does not account for any inherent bias in the laboratory's group of measurement procedures.

The last approach to bias for the purposes of QC planning is to assume bias is equal to zero. This approach recognizes that many measurement procedures trace their calibration to internationally recognized standards and that the calibration process should minimize actual bias. Although the actual bias may not be zero, for many measurement procedures it is small enough that it can be treated as zero. This approach recognizes that assessing the actual measurement procedure bias may not be practical and that using an estimate of relative bias may give a skewed perception of the actual bias that would not be useful in a QC plan. When bias is assumed to be zero, the QC plan is intended to identify deviations from a stable operating condition. The appropriate approach to bias is dependent on the technical limitations of assessing bias and the resources available to the individual laboratory.

3.3.2 Imprecision

A quantitative measurement procedure's imprecision is typically expressed as an SD or CV. In product inserts (instructions for use) and in the clinical chemistry literature describing an individual measurement procedure's performance, SDs and/or CVs are commonly encountered for two kinds of precision: **repeatability** (also called within-run precision) and **within-laboratory precision** (also called within-device, intermediate, or, formerly, total precision [see CLSI document EP05[24]]). Repeatability refers to variability over a short period of time, often less than 24 hours. Within-laboratory precision refers to variability over a longer period of time.

From a clinical point of view, repeatability is rarely of interest. Generally, within-laboratory precision estimates are clinically more relevant because they reflect variability over time intervals somewhat more representative of intervals between repeat measurements for a patient being monitored for a chronic disease or for response to treatment. Similarly for QC purposes, the within-laboratory precision is more relevant to a measurement procedure's stable, long-term performance that a QC strategy monitors.

Estimates of the within-laboratory SD for use in a QC strategy should be based on results from a long enough time period to adequately represent the types of influences that contribute to the measurement procedure's long-term, in-control imprecision. For example, contributions from electronic noise, pipette performance, detector performance, temperature control, daily recalibration, and similar sources are adequately represented in data from a modest time interval, such as a few weeks. However, contributions from periodic recalibration, changes in bottles of reagents, changes in lots of reagents or calibrators, maintenance procedures, and similar events that occur less frequently need much longer time intervals, such as several months or more, to be adequately represented in an estimate of the SD that reflects a measurement procedure's stable, long-term performance.

Product inserts typically report within-laboratory precision estimates based on the CLSI document EP05,[24] which has been devoted to precision evaluation since its first release in 1981. CLSI document EP05[24] includes a standardized protocol to estimate imprecision that is intentionally limited to a single-site precision study design that calls for measurements over as few as 20 days using a single lot of reagent and a single instrument. A similar protocol is also typical for most published studies. Such a study is usually completed in less than a month, but this time interval falls short of the clinically relevant time period for many measurement procedures.

SDs based on measurements obtained in less than a month are expected to underestimate the SDs that represent stable, long-term, in-control performance essential for effective statistical QC and for valid assessment of performance through Sigma metrics (see Subchapter 3.3.3). The time needed to achieve reliable representation of all important sources of variability depends on the measurement procedure. For

example, for a procedure calibrated every day, measurements over 20 days fairly reliably represent that source of variability, as well as other sources of variability that are exercised every day (eg, pipetting error). However, it may take over four months to achieve comparably reliable representation of calibration variability when the procedure calls for recalibration on a weekly basis. Similarly, contributions from other periodic or occasional sources of variability that may be important contributors to a measurement procedure's long-term performance need several months to be adequately represented.

Consequently, within-laboratory precision estimates provided in product inserts or in independent literature are generally regarded as, at best, lower bounds for the SD needed for an effective QC strategy or for evaluation of a measurement procedure's Sigma metric performance (see Subchapter 3.3.3), which is part of the information required for the QC strategy. Challenges include how to approximate the relevant SD in the short term for newly introduced measurement procedures and how to react to QC rule violations when the estimate of SD has not yet achieved adequate reliability. Subchapter 5.3.1 describes approaches that a laboratory can take to obtain an estimate of the SD that represents stable, long-term, in-control performance for a measurement procedure and includes all or most of the important sources of both short- and long-term influences on imprecision.

3.3.3 Sigma Metric

Historically, the integrated approach of combining bias and imprecision and comparing the resulting estimate of total error to the TEa has used an index relating estimated total analytical error to the quality requirement. The index was commonly referred to as process capability. More recently, for use in the medical laboratory, the index has been related to the quality monitoring concepts of Six Sigma and has been called the Sigma metric.[3,9,25] The Sigma metric is expressed numerically and is inversely related to the risk of failure of the measurement procedure. A high Sigma of six or higher represents an extremely low failure rate, while a low Sigma of three represents a much higher failure rate. Sigma metrics can be calculated for each measurement procedure and used to help guide laboratorians in designing an appropriate QC plan. In general, higher Sigma values translate to use of a less stringent QC strategy, and lower Sigma values indicate that a measurement procedure may require more QC to detect process failures. Subchapter 5.5.1 provides practical examples for selecting a QC strategy based on Sigma metric values.

The Sigma metric may be calculated using either repeatability or within-laboratory imprecision as the estimate of the SD. However, for the most useful estimates of the Sigma metric, the within-laboratory SD is the best choice. It should be estimated following the guidance in Subchapter 5.3.1, recognizing that SD determined over a period of months best characterizes long-term stable measurement procedure performance.

The Sigma metric at concentration *x* can be calculated as:

$$\text{Sigma}(x) = \frac{\text{TEa}(x) - |\text{Bias}(x)|}{\text{SD}(x)} \quad (1)$$

for which TEa(x), Bias(x), and SD(x) are the TEa, bias, and SD at concentration *x*.

If TEa is given as a percent, then:

$$\text{Sigma}(x) = \frac{\text{TEa}\% \bullet \frac{x}{100} - |\text{Bias}(x)|}{\text{SD}(x)} \quad (2)$$

A downside of the Sigma metric is its reliance on TEa and calculation of bias. The outcome of the calculation can change significantly from unacceptable performance to acceptable performance by merely selecting a different TEa or using a different assessment of bias. The other limitation of the Sigma metric

is that there is no single metric that characterizes measurement procedure performance over the entire measuring interval and, in many cases, for different medical uses of a given laboratory measurement procedure result. Rather, there are multiple Sigma metrics, each associated with a different measurand concentration or medical use of a laboratory measurement procedure. Consequently, there may be different Sigma metric values for each concentration of control material used or for each different medical use of a measurement procedure result. Despite these limitations, knowing Sigma at a given concentration may be beneficial because the laboratory can isolate QC requirements for a medical decision limit. When practical, basing the QC strategy on the most stringent Sigma performance metric is a conservative approach that minimizes the risk of harm for a patient.

3.4 Types of Out-of-Control Conditions

There are two basic classifications for out-of-control conditions: transient and persistent. Some examples of each are shown in Table 1. Transient conditions may affect a single sample or multiple samples over a short period of time. Due to the transient nature of these conditions, the condition may not be present at the next scheduled QC event and therefore not be detected.

Persistent out-of-control conditions continue until they are detected and the root cause eliminated. Persistent conditions fall into two categories: those conditions that alter the constant error or bias of the measurement procedure, and those conditions that alter the random error or imprecision of the measurement procedure.[26,27] Statistical QC strategies can detect changes to both bias and imprecision. Often, QC strategies are primarily focused on detecting changes to measuring system bias because bias often has a greater clinical effect. Subchapter 5.5.1 discusses selection of QC rules.

Out-of-control conditions that increase the bias of the measurement procedure usually appear as a change in the observed values of QC results compared to the stable target value. This change can occur abruptly over a short period of time, commonly referred to as a shift, or more gradually over a longer time, commonly referred to as a drift or trend.

Out-of-control conditions that cause a change in the random error of the measurement procedure usually appear as an increase in frequency of QC failures with both positive and negative differences from the stable target value. Changes in random error may be identified by an SD for a recent group of QC results that is larger than the stable SD. Generally, statistical QC is designed only to detect out-of-control random error conditions that cause a persistent increase in SD.

Table 1. Examples of Possible Causes for Out-of-Control Conditions

Possible Cause	Type of Error	Nature
Clot or debris in pipette	Systematic or random error	Transient or Persistent
Inadequate wash	Systematic or random error	Transient or Persistent
Incorrect liquid transfer volumes	Systematic or random error	Transient or Persistent
Temperature control	Systematic or random error	Persistent
Electronic noise	Systematic or random error	Transient or Persistent
Calibration problem	Systematic error	Persistent
Calibrator deterioration	Systematic error	Persistent
Reagent deterioration	Systematic error	Persistent
Deterioration of QC material	Systematic error	Persistent
Lack of calibration following major maintenance	Systematic error	Persistent

Abbreviation: QC, quality control.

3.5 Quality Control Rules

A QC strategy involves choosing which QC materials to measure and how many, when to schedule QC measurements, and which QC rule(s) to use to evaluate the QC results. A QC rule is a formal decision-making process that takes the results from one or more QC measurements and makes a decision either that the measurement procedure is performing in its stable in-control state (QC rule acceptance), or that the measurement procedure is not performing in its stable in-control state (QC rule rejection).

Chapter 4: Assessing Quality Control Performance

This chapter includes:

- Understanding false rejection rates
- Detecting out-of-control conditions
- Understanding the expected number of patient examinations before detecting an out-of-control condition

In general, QC performance assessment involves predicting various outcome measures for a given QC strategy during stable in-control operation and over a range of possible types and magnitudes of out-of-control conditions. During stable in-control operation, the primary outcome metrics of interest are:

- Probability of a QC false rejection
- Expected number of QC events between false rejections
- Expected number of patient examinations between false rejections
- Expected length of time between false rejections

When an out-of-control condition occurs, useful outcome metrics related to patient risk include:

- Probability the QC rule will detect the out-of-control condition
- Magnitude of an out-of-control condition that will be detected with a stated probability
- Expected number of QC events to detect the out-of-control condition
- Expected number of patient examinations affected by an out-of-control condition before detection
- Expected number of erroneous patient results reported before an out-of-control condition is detected

4.1 False Rejection Rate

4.1.1 Probability of False Rejection

When a laboratory's testing process is operating in its stable in-control state and a QC rule is evaluated, there is a chance that the QC rule will reject. This is referred to as a false rejection. The probability of false rejections depends on the number of QC concentrations examined, the total number of QC results evaluated, and the QC rule(s) used.[28] The probability of false rejection can be predicted either mathematically, by computer simulation, or by empirical evaluation of retrospective laboratory data. It is desirable to have the probability of false rejection as low as possible. However, in many cases, lowering the false rejection rate also lowers the ability to detect out-of-control conditions when they occur, so there is always a need to balance the desire for a low false rejection rate with the required error detection capability.

4.1.2 Expected Number of Quality Control Events Between False Rejections

A quantity closely related to the probability of false rejection is the expected number of QC events between false rejections.[28] In many situations, there is an inverse relationship between the probability of a QC rule rejection and the expected number of QC events before a QC rule rejection. For example, if the probability of false rejection is 0.01 (1 in 100), then the expected number of QC events between false rejections is 100.

4.1.3 Expected Number of Patient Examinations Between False Rejections

The probability of false rejection is the probability of a QC rule rejection when the measurement procedure is performing in a stable condition. The rate of false rejections not only depends on the probability of a false rejection, but also on how often QC rules are evaluated. The rate of false rejections are characterized in terms of number of patient specimens tested between false rejections or in terms of elapsed time between false rejections. Both metrics have value depending on the situation. The average number of specimens between false rejections depends on the average number of specimens measured between QC events and the expected number of QC events between false rejections.[29]

4.1.4 Expected Length of Time Between False Rejections

The average (expected) length of time between false rejections depends on the length of time between QC events and the expected number of QC events between false rejections. The shorter the time interval between QC events and/or the fewer the expected number of QC events between false rejections for the QC rule, the shorter the average length of time between false rejections.

4.2 Detection of Out-of-Control Conditions

4.2.1 Probability of Detecting an Out-of-Control Condition

When a laboratory's testing process experiences an out-of-control condition and a QC rule is evaluated, there is a chance that the QC rule will give a rejection. This chance is called the probability of error detection.[28] In general, for small out-of-control conditions the probability of error detection is low, and for large out-of-control error conditions the probability of error detection is high. In other words, the larger the out-of-control condition, the more likely it is detected.

It is desirable to detect an out-of-control condition of a magnitude associated with an increased probability of producing erroneous patient results not fit for their intended use as soon as possible. What is considered an unacceptably high probability of producing erroneous patient results depends on the likelihood of patient harm from an erroneous result, or the severity of patient harm due to the decisions made (action or inaction) based on an erroneous result.

The probability of error detection can be predicted either mathematically, by computer simulation over a range of possible out-of-control conditions, or by empirical evaluation of retrospective laboratory data.

4.2.2 Expected Number of Quality Control Events Before Detecting an Out-of-Control Condition

An alternative quantity that is closely related to the probability of error detection is the expected (or average) number of QC events required to detect an out-of-control condition.[28] In general, for small out-of-control conditions, the expected number of QC events before a QC rule violation is high. Conversely, for large out-of-control error conditions, the expected number of QC events before a QC rule violation should be low. There is an inverse relationship between the probability of error detection and the expected number of QC events before an error detection; the higher the probability of error detection, the lower the expected number of QC events before detection. The expected number of QC events until error detection over a range of possible out-of-control conditions can be predicted either mathematically, by computer simulation, or by empirical evaluation of retrospective laboratory data.

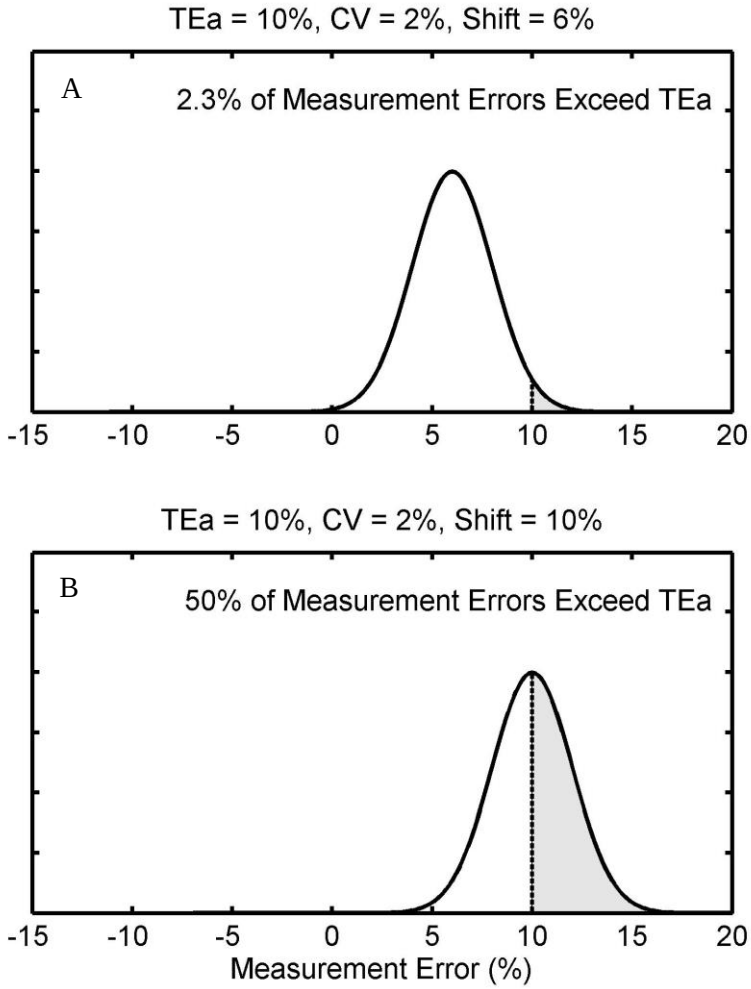
4.2.3 Expected Number of Patient Examinations Before Detecting an Out-of-Control Condition

The probability of detecting an out-of-control condition is the probability of a QC rule rejection when the QC results are evaluated in the presence of an out-of-control condition. However, the number of patients affected by an out-of-control condition not only depends on the probability of detecting an out-of-control

condition when a QC rule is evaluated but also on how frequently QC events are scheduled.[28] The more patient specimens tested between QC events, the larger the number of patient results potentially affected by an out-of-control condition before it is detected.

4.2.4 Expected Number of Erroneous Patient Results Before Detecting an Out-of-Control Condition

Not all patient results potentially affected by an out-of-control condition necessarily contain a measurement error large enough to make them unfit for their intended use. The percentage of affected patient results that contain an unacceptable measurement error depends on the magnitude of the out-of-control condition and when the error condition occurred. For example, if the quality requirement is that measurement error should not exceed 10%, and a measurement procedure with a 2% CV experiences an out-of-control shift of 6%, then all patient results examined during the existence of the 6% shift are affected by the shift, but only about 2.3% of the affected patient results are predicted to contain a measurement error exceeding 10%. Alternatively, if an out-of-control condition caused a 10% shift in the process, then 50% of the affected patient results are predicted to contain a measurement error exceeding 10% (see Figures 3A and 3B).



Abbreviations: CV, coefficient of variation; TEa, allowable total error.

Figures 3A and 3B. Influence of the Magnitude of a Change in Bias for a Measurement Procedure on the Number of Patient Results Affected by the Error Condition. The shaded areas represent the fraction with error that exceeds TEa.

The predicted number of patient results with unacceptably high measurement error are divided into three categories (see Table 2).

Table 2. Categories of Erroneous Patient Results

Category	Effect
Erroneous patient results are produced but are **not** reported before QC detects the out-of-control error condition.	There is no opportunity to create a hazardous situation.
Erroneous patient results are reported, but subsequent QC detects the out-of-control error condition soon enough to give the laboratory an opportunity to correct the erroneous results before they are acted upon.	A hazardous situation is avoided.
Erroneous patient results are reported and acted upon before the out-of-control error condition is detected.	A hazardous situation is created.

Abbreviation: QC, quality control.

The first two categories are the expected number of unreliable correctable results. The third category is the expected number of unreliable final results.

Chapter 5: Planning a Statistical Quality Control Strategy

This chapter includes:

- Defining the quality requirements goals for a measurement procedure
- Selecting the appropriate control materials and the levels for these materials
- Setting goals for QC performance
- Selecting a QC strategy based upon the QC performance
- Choosing an appropriate QC rule
- Choosing an appropriate QC schedule

5.1 Define the Quality Requirements

5.1.1 Sources of Information

In 1999, an international consensus conference proposed a hierarchy of sources of information and approaches for quality specifications for laboratory medicine.[30,31] The recommendations from this conference were updated in 2014.[32,33] The strategic conference described three models of goal setting listed in descending order of preference. Some approaches are better suited for certain measurands than for others. Irrespective of which approach is used, the laboratory director should consult with medical care providers to agree on an appropriate TEa for the patient population served.

5.1.1.1 Goals Based on Outcome Studies (Model 1)

Setting a TEa goal based on the effect of analytical performance of the measurement procedure on the clinical outcome is the preferred model. Outcome studies can be the direct assessment of either clinical outcomes for a group of patients or of "indirect" outcomes for which consequences of analytical performance on classifications or decisions regarding disease or risk for disease are investigated and related to the probability of patient outcomes.[34-36] Indirect outcome studies are often used to set TEa in laboratory practice guidelines. In an outcomes-based approach, the goals are relevant to patient care requirements. The disadvantage with this model is that it requires a close relationship between the measurand, medical decision making, and clinical outcomes that is only applicable to a relatively small number of measurands.

5.1.1.2 Goals Based on Biological Variability of the Measurand (Model 2)

Another concept for defining performance goals is that analytical error should be smaller than the natural biological variation for a given measurand. In this model, TEa is based on a fraction of the within- and between-individual biological variations of the measurand.[37,38] This model assumes that a small ratio between analytical error and expected biological variation will identify measurement procedure performance that relates to the medical requirements. Strengths of the biological variability approach are that it uses a defined statistical approach based on measurable biological variability parameters and that data on biological variability are available for many measurands.[39]

Weaknesses of this model are its lack of focus on clinical outcomes or medical requirements, that the difference in concentration to discern between healthy and diseased conditions is not considered, and that the reliability of some available biological variability data has been questioned.[40-44] In addition, for some measurands the biological variability cannot be measured for nondiseased persons; eg, serum human chorionic gonadotropin for nonpregnant women or for nonmalignant conditions. There is also the challenge that current technology may not be able to produce measurement procedures capable of achieving the biological variability–based goals for some measurands that are closely regulated

biologically. For these measurands, biological variability–based goals may represent aspirational goals for new technology development, but they may not be practical goals for currently available technology.

5.1.1.3 Goals Based on the State of the Art (Model 3)

In this approach, measurement procedure performance that represents the best that can be achieved by current technology, and/or is similar to that of peers, is defined as acceptable. An advantage of this model is that the information is accessible from internal QC data or from some PT/EQA surveys when commutable samples are used. A weakness of this approach is that PT/EQA samples are frequently not commutable with clinical patient specimens and large differences may be seen in PT/EQA schemes due to matrix-related errors that do not reflect the differences observed for patient specimens.[23] Model 3 also makes no assessment of the possible differences in clinical interpretation that could result from the differences observed in measured results.

5.1.2 Considerations in Setting Goals

There are no universally accepted TEa goals for measurands. It is likely that no single approach among those described is optimal for setting the TEa goals for all of the measurement procedures used in the laboratory. Therefore, the laboratory director should use the approach for each measurement procedure deemed most suitable for the laboratory's specific needs. If a TEa goal is selected that cannot be achieved by the current measurement procedure, then a new procedure should be considered. However, it is possible that no commercially available measurement procedure may be able to achieve the desired goal. When that is the case, it may be necessary to re-evaluate the desired goal or to use an appropriate QC strategy to identify a relatively small deterioration in measurement procedure performance to minimize risk of erroneous results being reported.

5.2 Select Control Materials

Control materials should have characteristics that enable them to provide information about the performance of a measurement procedure when making measurements with the intended patient specimen types. Ideally, the matrix of a QC material (eg, serum, urine, whole blood) should be the same as that of the patient specimens that are measured. However, the matrix is typically modified from that of a patient specimen because of the need for stabilizing agents, added measurands to achieve desired concentrations, and other manipulations associated with manufacturing QC materials. For those control materials that have a nonhuman or chemically contrived matrix to resemble human matrixes, the ability to make inferences about errors in patient specimens may be compromised.[7]

The matrix should be generally similar to that of the patient specimens; for example, a serum-based QC material is appropriate when the patient specimen is serum. However, it is not always practical to have an array of different QC material matrixes when the same measurement procedure is used to measure, for example, serum, plasma, urine, and cerebrospinal fluid specimens. The primary purpose of a QC material is to determine that a measurement procedure is performing as expected in order to confirm that the results for patient specimens are suitable for use in providing medical care. When the same measuring interval is used for different patient specimen matrixes, QC samples of a single matrix, with suitable concentrations, may be sufficient to monitor the performance. In the situation when a patient specimen matrix requires a different measuring interval than used for other patient specimens, it is necessary to ensure that a QC sample with matrix and concentration suitable for that measuring interval is included in the QC strategy.

A laboratory should obtain enough homogeneous and stable control material to last for an extended time interval, such as one or more years, when possible. Using the same lot of QC material optimizes the ability to establish expected results and evaluation criteria and to use the QC results to monitor the stability of a measurement procedure. In addition, the longer the same lot of QC material is used, the less

frequent is the need to establish baseline statistical characteristics for new lots of QC material. Vial-to-vial variability of the QC material should be much less than the variation expected for the measurement procedure being monitored. Open vial stability for claimed measurands in a QC material should meet the needs of the laboratory and be verified.

There are different types of control materials available to laboratories. Each has strengths and weaknesses. The types include:

- Control materials made and supplied by the manufacturer of the measurement procedure
- Control materials that are made by a third party for the manufacturer of the measurement procedure
- Control materials that are made by a third party and have no relationship to the measurement procedure manufacturer or to the calibrator used for the measurement procedure
- Patient specimen pools or other laboratory-prepared materials

If there is no appropriate QC material available, and laboratory-prepared materials are not practical or technically feasible, the approach to QC recommended in this guideline is not applicable.

5.2.1 Control Materials Made and Supplied by the Manufacturer of the Measurement Procedure

Control materials made and supplied by the instrument or reagent manufacturer are sometimes referred to as "kit" or "in-kit" controls. Such controls may be single measurand controls or have multiple measurands per vial. These controls may be manufactured from the same raw materials and based on the same or similar formulations as the calibrator set for the measuring system. They may be optimized to work on a specific instrument and/or with a specific reagent(s) and often do not work on other instruments or with other manufacturers' reagents. Optimized controls, especially if they mimic the calibrator, may not be able to detect some systematic errors.

5.2.2 Control Materials Made by a Third Party for the Manufacturer of the Measurement Procedure

Control materials made by a third party for the manufacturer of the measurement procedure are typically manufactured under contract for an instrument or reagent manufacturer. These controls are made to a specific formulation supplied by the instrument or reagent manufacturer and are supplied to laboratories either by the instrument or reagent manufacturer or directly from the third party manufacturer that made them. They may have a formulation similar to the manufacturer's calibrators. As noted in Subchapter 5.2.1, if the formulation is too similar to calibrators and/or too dissimilar to patient specimens, some changes in method performance may not be detected as effectively.

5.2.3 Control Materials Made by a Third Party That Has No Relationship to the Measurement Procedure Manufacturer or to the Calibrator Used for the Measurement Procedure

Control materials made by a third party that has no relationship to the measurement procedure manufacturer or to the calibrator, sometimes referred to as third party controls, are developed independently without any influence from the instrument or reagent manufacturer. The control materials are independent of any specific instrument, calibrator, or reagent set. Such control materials can typically be used across multiple measuring systems. These types of control materials are most often made from a human matrix such as serum, blood, plasma, or urine. The matrixes may be modified to meet laboratory expectations for stability or to achieve required concentration values. Consequently, such control materials may exhibit matrix effects of varying magnitudes when used with analytical measurement procedures that are matrix sensitive.

5.2.4 Patient Specimen Pools or Other Laboratory-Prepared Materials

The laboratory can prepare and aliquot pools of patient specimens or prepare other suitable samples for use as controls. It may be necessary to supplement pooled samples with purified analytes to obtain concentrations suitable for QC monitoring. Note that pooling and supplementing can alter the matrix of the material, which can affect its usefulness. In addition, it may be difficult to achieve clinically relevant concentrations that challenge the measuring interval of a measurement procedure. The stability of a pool of patient specimens can be a limitation for some measurands.

5.2.5 Relation to Calibrators

QC materials should be different from the calibrator materials to ensure that the QC results provide an independent assessment of the measurement procedure's performance in its entirety, including the procedure for calibration. If it is necessary to use calibrators as QC materials, the lot number used for calibration needs to be different from the lot number used for QC.

5.2.6 Concentrations of Measurands in Control Materials

The number of concentrations of QC materials should be sufficient to determine acceptable method performance over the measuring interval of interest. The appropriate concentrations at which to monitor a procedure's performance are based on both clinical decision values and the analytical measuring interval. Ideally, at least one control material should contain a concentration of the measurand at or near the clinical decision level. Accreditor requirements and government regulations may specify a minimum number of control concentrations for certain laboratory measurement procedures.

5.2.6.1 Clinical Decision Values

Measurand concentrations at clinically relevant values are appropriate for monitoring performance and for providing documentation of the suitability of results. The imprecision data from QC results are also useful for assessing agreement among different measurement procedures (see CLSI document EP31[45]) or for verifying performance when changing reagent lots (see CLSI document EP26[46]), both of which are important at clinical decision concentrations.

5.2.6.2 Analytical Decision Values

Measurand concentrations at levels dictated by the analytical performance characteristics are also appropriate for monitoring measurement procedures. For example, performance near the lower or upper limits of the measuring interval may be important for verifying that the measuring system remains stable over the entire measuring interval. There may be practical limitations in the availability of QC materials with concentrations that cover the entire measuring interval, in which case alternative approaches to verifying the measuring interval are needed (see CLSI document EP06[47]).

5.2.6.3 Number of Quality Control Concentrations

For most measurement procedures, a minimum of two concentrations of QC materials is recommended. Using QC samples at more concentrations may be necessary to adequately monitor measurement procedure performance and to enable application of QC rules that improve detection and interpretation of potential measurement errors (eg, proportional vs constant, random vs systematic). Note that for some measurands, the clinically relevant concentrations may span a large part of the measuring interval, and three or more concentrations may be needed for adequate performance monitoring. For example, newborn and adult concentrations of bilirubin are very different, and at least three concentrations may be needed for adequately verifying measurement procedure performance.

5.2.7 Quality Control Concentrations for Quantitative Measurements Reported as Qualitative Values

When quantitative measurements are transformed to qualitative results based on a threshold value that determines a negative or positive response, analogous approaches to QC are applicable. In this situation, two QC concentrations are needed: one below and one above the threshold value. The magnitude of the differences from the threshold value should be chosen so the QC values monitor performance over the restricted measuring interval around the threshold value. The quantitative signal, or concentration result, is used as the QC value and its acceptability is evaluated using the same assessment rules used for a quantitative reported value.

5.2.8 Lyophilized and Liquid Controls

Control materials are generally stabilized in order to have a long useful life. There are two common approaches to stabilize QC material: lyophilization or "frozen" liquid. Lyophilized materials are the most stable form in which control materials are supplied. They are characterized by long shelf life. They need reconstitution with a specified diluent. The reconstitution process adds some degree of variability to the estimate of imprecision based on the QC results. Also, it is necessary to allow enough time for the material to fully reconstitute before use. These materials are ideal for laboratories in locations where freezers are uncommon or expensive to run, or in laboratories that have limited freezer space.

Liquid materials provide convenience but typically need frozen storage. There is no reconstitution needed, so the estimate of imprecision based on the QC results may be more representative of the measurement procedure imprecision. However, frozen liquid controls need careful mixing before use, and the stabilizing agents may interfere with or contribute to imprecision estimates for some measurement procedures.

5.3 Determine Target Values and Standard Deviations for Quality Control Materials That Represent Stable Analytical Performance

A target value and SD for a particular control material are established by the laboratory. The mean, used as the target value, and SD of results are established by repeated measurements of the QC materials by the measurement procedure used by the laboratory. Control limits are then calculated from the target value and SD observed in the laboratory when the measurement procedure is operating in a stable condition. When control materials are accompanied by a product insert with assigned values provided by the manufacturer, these insert values should be used only as guides and not as a replacement for target values and SDs established by the laboratory.

5.3.1 Stable Total Imprecision (Standard Deviation) for Each Control Material

When there is a history of QC data from an extended period of stable operation of the measurement procedure, the established estimate of the SD (or CV) can be used with a new lot of control material. Imprecision is a characteristic of a measurement procedure and is generally the same irrespective of the lot of QC material used. There may be exceptions for a new formulation of a QC material, in which case the SD can be updated after sufficient experience is obtained with the new lot. The established SD is appropriate when the new lot of QC material has a similar target concentration for the measurand of interest as for the previous lot. If the target concentration for the new lot differs enough from the previous lot so that use of an established SD is not appropriate, then a new SD can be estimated using the established CV at the closest prior lot concentration as long as the CV is approximately constant over the concentration interval involved. The SD for the new lot is estimated by multiplying the estimated mean for the new lot of QC material times the CV (%), divided by 100. (This is referred to as the "simple formula" for sample SD calculation.)

When replacing an existing measurement procedure with a new one in the laboratory, it is often possible to use the existing measurement procedure's SD as an initial estimate for the new one. When doing so, assumptions are made that the existing measurement procedure's SD is suitable to confirm that the results are appropriate for medical use, and that the new measurement procedure performs similarly or better based on its validation data. Once enough QC results have been accumulated for the new measurement procedure, the initial SD should be updated to reflect the long-term variability of the new measurement procedure.

When historic estimates are not available, initial estimates of SD are obtained by measuring at least 20 data points on separate days. The measurements obtained in this initial value assignment study should represent the measurement procedure in its stable in-control state. Conditions for the study should mimic routine operation as closely as possible. For example, if an opened bottle of QC material is used for more than one day during routine operation, the same practice should occur during the initial SD estimation stage so that QC material stability is reflected in the initial estimate of the SD. A Levey-Jennings plot (see Appendix A) should be constructed from the measurements for each control material being evaluated. Visual inspection of the Levey-Jennings plot may identify a pronounced drift or shift in the results over time, or an occasional highly deviant result. If no pronounced drifts, shifts, or outliers are seen, then the laboratory can use the 20 data point estimates of SD to proceed with monitoring the measurement procedure during routine operation until improved estimates based on a larger sample set can be obtained. If only a single data point is collected each day, then the SD can be reasonably estimated using the simple formula for the sample SD. If more than one data point per day is obtained, then the simple formula tends to underestimate the long-term SD but in most cases still provides an adequate estimate. Alternatively, the SD can be estimated using a one-way analysis of variance approach such as the approach described in CLSI document EP15.[48]

During the initial phase of routine operation using an initial estimate of SD, the laboratory should monitor its QC data as they are accumulated over time. Because of the limited reliability of the initial SD estimates, evaluation and response to QC rule violations should consider the possibility that the SD limits are inadequately estimated. Computing the cumulative SD over the first several months of operation gives a better estimate of the SD because additional components of longer-term sources of variability are included in the data. Long-term sources of variability are, for example, different calibration cycles, different reagent bottles or lots, preventive maintenance, component replacement, and environmental factors.

For some measurement procedures, QC materials may exhibit a change in numeric values when a reagent lot is changed (see CLSI document EP26[46]). The shift in values is caused by a change in the matrix-related interaction of the QC material with a specific reagent lot. Such a change in values is an artifact of the interaction of the QC material and a specific reagent lot for that measurement procedure. Note that the SD of the measurement procedure is unlikely to be affected by a reagent lot change. However, the cumulative SD is inflated by the artifactual shift in values if QC data obtained with different reagent lots are included in the calculation, and the estimated cumulative SD is not representative of the SD expected when measuring patient specimens. Consequently, for measurement procedures in which this artifact can occur, the SD should be estimated using data from a single reagent lot. Alternatively, when QC data from more than one reagent lot is needed to provide an adequate time interval to include the important sources of long-term variability, the pooled SD calculation described in equation (3) should be used.

Equation (3) is used to combine (pool) QC results from more than one time period. This pooling equation may be used when results from more than one reagent lot or QC material lot are combined (eg, due to short stability of the QC material) to provide an adequate time interval for including the important sources of long-term variability. SD_i is the SD for the *i*th time interval of stable performance, and n_i is the number of QC results obtained during the interval. If *k* time intervals of stable performance are available, then a pooled SD is estimated as:

$$\mathrm{SD}_{\mathrm{pooled}} = \sqrt{\frac{(n_1-1)\mathrm{SD}_1^2+(n_2-1)\mathrm{SD}_2^2+\cdots+(n_k-1)\mathrm{SD}_k^2}{n_1+n_2+\cdots+n_k-k}} \quad (3)$$

If pooling across multiple stable time intervals is not possible or appropriate, then the limitations of a less reliable SD may have to be accommodated in the QC plan.

5.3.2 Target Value for Each Control Material

If there is no history of QC data, the mean can be estimated from the data points used to estimate the SD and used as the initial target value (see Subchapter 5.3.1).

If there is a QC material in current use, and an estimate of the SD is not needed for the new QC lot, then a new lot of QC material should be analyzed for each measurand of interest in parallel with the lot of control material in current use. In most cases, 10 measurements made on separate days are adequate to estimate a mean that is suitable for use as the initial target value. A minimum of 10 days enables some day-to-day sources of variability in the measurement procedure to be reasonably represented in the mean value. Periodically computing the cumulative mean over the first several months of operation gives a better estimate of the mean because additional components of longer-term sources of variability are included in the data. Several calibration events during the time interval used to establish the target value for a new lot of QC material should be included. Also note that when an opened bottle of QC material is used for more than one day, the same bottle should be used for the number of days of intended use to allow measurand stability to be reflected in the mean value.

There are situations in which the laboratory needs to more quickly establish a target value for a new lot of QC material. In such cases, the mean from fewer days' measurements may be used, including more than one measurement per day. A target value so established is considered temporary and should be updated as soon as sufficient data are obtained to estimate a stable mean.

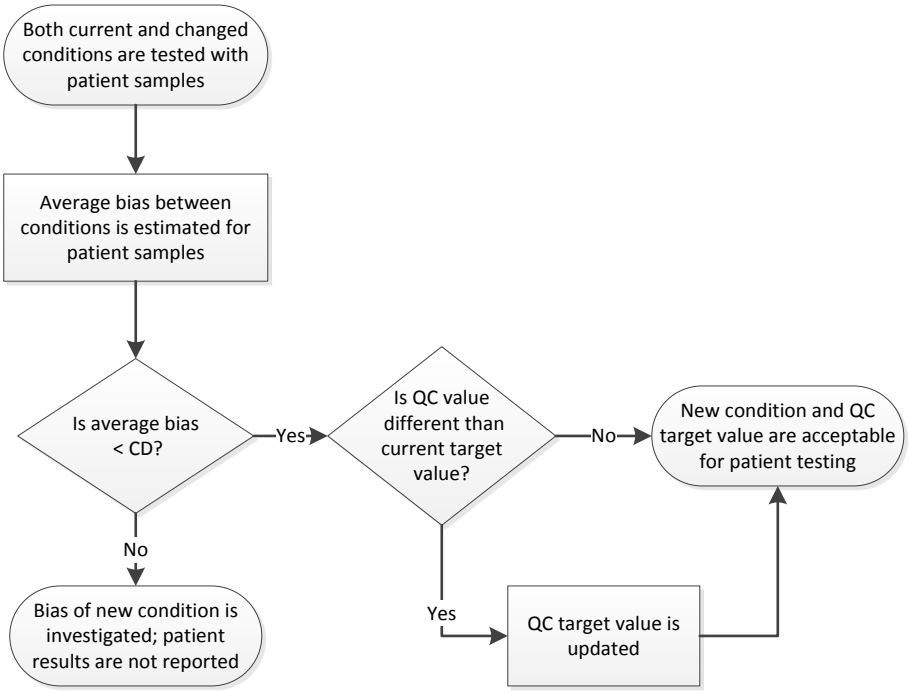
5.3.2.1 Adjusting the Target Value During the Life of the Lot of Quality Control Material

In principle, the target value should be established and then not changed in order to allow the performance of a measurement procedure to be tracked over time. However, the expected value for a QC material can be influenced by changes in measurement conditions that may not affect patient results, such as reagent lot changes,[49] some maintenance procedures, or deterioration of a measurand during the expected shelf life of a QC material. Because there is no change for patient results, a change in a QC value does not represent any problem with measurement procedure performance. Such changes in QC results are artifacts of the interaction of the reagent with the altered sample matrix of the QC material or of changes in measurement procedure components that are sometimes unidentifiable. When this situation occurs, it is necessary to update the QC target values to reflect the changed performance of the QC material. Failure to update the target value when needed introduces an artifactual bias that negatively affects the ability of the QC acceptance criteria to identify erroneous measurement conditions. For example, a target value that is incorrectly high causes measurement error conditions that produce high values to be poorly identified and the expected distribution of lower values to cause false QC alerts.

Guidance on verifying performance for patient results following reagent lot changes is available in CLSI document EP26[46] and is also applicable to verifying the continuing acceptable performance for patient results following any type of change in measurement conditions that may alter the target value for a QC material without affecting the results for patient specimens. Figure 4 illustrates the general approach to evaluate the suitability of a QC target value following any change in conditions that can alter the QC target value without affecting the results for patient specimens.

NOTE: Local regulatory requirements may preclude the laboratory from updating QC target values when a shift in QC results is noted. In this situation, all local regulatory requirements need to be followed, and

the laboratory should contact the manufacturer(s) of the measurement procedure and QC material for assistance.



* Five basic symbols are used in process flow charts: Oval (signifies the beginning or end of a process), Arrow (connects process activities), Box (designates process activities), Diamond (includes a question with alternative "Yes" and "No" responses), Pentagon (signifies another process).
Abbreviations: CD, critical difference that would alter a decision made for patient care; QC, quality control.

Figure 4. Flow Chart for Verifying Performance Following Any Change in Conditions That Alters the QC Target Value Without Affecting the Results for Patient Specimens.* A change in conditions could be a new reagent lot, a component replacement, maintenance, or other procedure that may affect the QC differently than it affects the patient results.

If the laboratory has specific concerns about a particular QC material, measurement procedure, or reagent lot, it can use the verification techniques described in CLSI document EP09,[21] which uses 40 patient specimens to more robustly estimate the difference in patient results between two reagent lots or other measurement conditions. In this type of study, the comparative measurement procedure is defined as the current reagent lot or measurement condition before a change in component, and the candidate measurement procedure is defined as the new reagent lot or the measurement condition after a change in component.

Shifts or changes in QC values can occur independently of a change in reagent lot or measurement procedure component, and it is not always possible to compare patient specimens tested before and after the change occurred, especially when the time interval to identify the change exceeds more than a few days. It is not appropriate to adjust QC target values until the laboratory can confirm there were no changes in patient results, or a robust troubleshooting and investigation of the measurement procedure has failed to identify an assignable cause for the shift in a QC value. Appendix B provides a checklist for laboratorians to use when investigating QC shifts and provides an example checklist for documenting the investigation.

5.3.2.2 Cumulative Mean Values May Be Inappropriate as Target Values for a Quality Control Material

The cumulative (initial use to date) mean of QC results stabilizes over time as the number of values increases and additional sources of variability in individual measurements are included in the data. After a

period of time sufficient to include most sources of measurement variability, the cumulative mean may be a good estimate for a stable target value. However, the expected value for a QC material can be influenced by changes in measurement conditions that may not affect patient results, such as reagent lot changes or some maintenance procedures. In these situations, the cumulative mean is not appropriate as a target value, and it can take considerable time for the cumulative mean to reflect the altered measurement conditions. Consequently, cumulative means may be inappropriate as target values, and it is preferable to update the target value, as described in Subchapter 5.3.2.1.

5.3.3 Using Assayed Control Materials

Some control materials have an assigned target value and a range of acceptable values in the product labeling. The product insert should be examined for the intended use of such QC materials.

For example, when assayed QC materials are provided by the manufacturer of a measurement procedure, they may be intended to determine if that procedure meets the manufacturer's specifications and may be suitable for use by a laboratory.

If assayed QC materials are provided by the manufacturer for use with several different measurement procedures, caution is needed because the nominal target values and acceptable ranges may not be suitable for different measurement procedures. Because of the general limitation of matrix-related bias with different reagent lots and among different measurement procedures, and limitations in the number of replicates and number of representative measurement procedures used for the estimation of statistical parameters, the target value and SD may be suitable as general information but are unlikely to be suitable for QC of a particular measurement procedure in a single laboratory. By the time a product insert is published and a control material is released for sale, the assigned values may or may not have continued relevance. To complicate matters, the product insert typically gives no indication of how the values were actually derived. Consequently, the laboratory should calculate the target value and SD that reflects performance of the measurement procedure in their laboratory environment.

5.4 Set Goals for Quality Control Performance

The goals (criteria) for acceptable QC results are primarily based on the performance that a measurement procedure is capable of achieving because the purpose for making the QC measurements is to verify that a measurement procedure continues to meet its expected analytical performance. However, less stringent QC acceptance criteria may be used if the risk of harm to a patient is kept at acceptable levels.

5.4.1 Influence of Likelihood of Patient Harm on Quality Control Performance Goals

An erroneous laboratory result is a hazardous condition that may cause harm if acted on. Harm can be caused by performing a clinical intervention that is not appropriate, or by failing to initiate a clinical intervention that is needed to prevent harm. The laboratory's tolerance for reporting erroneous results should depend on the likelihood that an erroneous result will cause patient harm and on the severity of the patient harm. The more likely or more severe the patient harm, the more frequently QC events should be scheduled and the more powerful the laboratory's QC rules should be in order to effectively limit the number of erroneous results reported in the event of an out-of-control condition.

5.4.2 Quality Control Performance Goals Cannot Alter Measurement Procedure Performance

It is important to note that setting QC rule acceptance criteria does not change the performance of a measurement procedure. QC results that do not meet QC rule acceptance criteria are intended to indicate that a change in performance of a measurement procedure has occurred. If improved measurement procedure performance is desired or the performance is barely adequate to meet clinical needs, the performance cannot be improved by more stringent QC rule acceptance criteria. However, more stringent

acceptance criteria can detect smaller deviations in performance, but there will be an increased rate of QC rule failures causing an increased amount of troubleshooting, repeated measurements for patient specimens, and likely delays in reporting patient results. The additional work involved in following up on these QC rule failures reduces efficiency and increases operational costs that may be necessary when a measurement procedure's stable performance is similar to or greater than the TEa.

If the inadequacy of a measurement procedure is due to systematic drift or shift, more frequent QC events with acceptance criteria consistent with stable performance may identify such conditions earlier. However, if the inadequacy is caused by imprecision, more stringent acceptance criteria is of no benefit because the influence of imprecision on any individual measurement affects QC results and patient results randomly. Consequently, an observation that a QC result is within a more stringent acceptance criteria does not predict that a patient result will be consistent with the same imprecision criteria.

5.5 Select a Quality Control Strategy Based on Performance Goals

The QC strategy (QC concentrations, number of concentrations, number of measurements at each QC concentration, QC rules, and QC schedule) selected by the laboratory should be based on the quality requirements (see Subchapters 3.2 and 5.1) and specified patient risk–based performance goals (see Subchapter 3.1). For a candidate QC rule and number of QC results evaluated, the probability of false rejection (see Subchapter 4.1.1), the expected number of QC events between false rejections (see Subchapter 4.1.2), the probability of detecting an out-of-control condition (see Subchapter 4.2.1), and the expected number of QC events before detecting an out-of-control condition (see Subchapter 4.2.2) can be computed mathematically (for some QC rules) or by computer simulation. Power functions giving probabilities of false rejection and error detection have appeared in the laboratory medicine literature for many common QC rules. The computations need advanced computer software and assume the imprecision is described by a gaussian (normal) distribution.

Likewise, for a candidate QC rule, number of QC results evaluated, and QC schedule, the expected number of patient examinations between false rejections (see Subchapter 4.1.3), the expected time between false rejections (see Subchapter 4.1.4), the expected number of patient examinations before error detection (see Subchapter 4.2.3), and the expected number of erroneous patient results before error detection (see Subchapter 4.2.4) can be computed mathematically or by computer simulation.

In the absence of advanced computer software that can predict the performance of a specific candidate QC strategy, general guidelines are helpful for selecting an appropriate QC strategy that meets performance goals based on patient risk. For decisions about when to schedule QC events, Subchapter 5.4 provides some valuable general guidance.

5.5.1 Quality Control Rules

All statistical QC rules use one or more measured values obtained from QC samples to make a decision about whether the measurement procedure is operating in its stable in-control state. A wide variety of QC rules has been proposed for use in the medical laboratory.[28,50,51]

Many of the QC rules that have appeared in the laboratory medicine literature are "counting" rules. The decision criteria are based on counting the number of QC results that violate a specified control limit. These types of counting rules can be represented by abbreviations of the form A_L, for which "A" represents the number of control observations and "L" is a control limit. If "A" or more control observations exceed the rule's control limits, then the decision is made that the measurement procedure is not in control. For example, 1_{3S} refers to a control rule to assess whether a single control result is beyond three SDs from the target value. Similarly, 2_{2S} refers to a control rule to assess whether results from two

control samples both exceed two SDs from the target value in the same direction. Examples of QC counting rules are:

- **1_{3S} rule:** Reject if any QC result from the current QC event is more than three SDs from the QC target values.

- **$1_{3.5S}$ rule:** Reject if any QC result from the current QC event is more than 3.5 SDs from the QC target values.

- **2_{2S} rule:** Reject if two QC results exceed 2 SDs from the QC rule's target values in the same direction. This rule can be applied to QC results from the same control material obtained from two successive QC events (within QC concentrations), to the QC results from two different QC materials measured in the current QC event (across QC concentrations), or both.

- **2 of 3_{2S} rule:** Reject if two out of three QC results exceed 2 SDs from the QC rule's target values in the same direction. This rule can be applied within QC concentrations, across QC concentrations, or both.

- **3_{1S} rule:** Reject if three QC results exceed 1 SD from the QC target values in the same direction. This rule can be applied within QC concentrations, across QC concentrations, or both.

- **4_{1S} rule:** Reject if four QC results exceed 1 SD from the QC target values in the same direction. This rule can be applied within QC concentrations, across QC concentrations, or both.

- **10_{1S} rule:** Reject if 10 QC results exceed 1 SD from the QC target values in the same direction. This rule can be applied within QC concentrations, across QC concentrations, or both.

NOTE 1: The 3_{1S} and 4_{1S} rules are generally applied across QC concentrations. The 10_{1S} rule is generally applied to single concentrations.

NOTE 2: Counting rules for which the count is a multiple of 2, such as the 2_{2S} and 4_{1S} rules, are generally used when two concentration levels of QC are evaluated. Rules for which the count is a multiple of 3, such as the 2 of 3_{2S} and 3_{1S} rules, are generally used when three concentration levels of QC are evaluated.

Another general class of QC rules combines multiple QC results into a single value that is compared to a specified decision limit in order to decide whether the measurement procedure is in or out of control. In order to combine QC results from different concentrations of control material, the results are transformed by subtracting the QC target value from the measured value and dividing by the SD for the QC concentration. These transformed values are sometimes referred to as z-scores or standard deviation intervals (SDIs). Examples of QC rules of this type are:

- **Mean rule:** Reject if the absolute value of the mean of z-scores or SDIs obtained from the QC results in the current QC event exceeds a specified control limit.

- **Moving mean rule:** Reject if the absolute value of the mean of z-scores or SDIs of the most recent *N* QC results exceeds a specified control limit, for which *N* is the number of QC results to include in the mean calculation.

- **Exponentially weighted moving average (EWMA) rule:** Reject if the absolute value of the EWMA of current and previous QC results (z-score or SDIs) exceeds a specified control limit. The calculation of the EWMA depends on a weighting constant that is between zero and one.[52,53]

- **Cumulative sum (CUSUM) rule:** Reject if the CUSUM of consecutive QC results (z-scores or SDIs) exceeding a defined threshold level exceeds a specified control limit.[54]

- **Range rule:** Reject if the range of z-scores or SDIs of the QC results from the current QC event exceeds a specified control limit.

The control limits for the mean and range rule are typically set so that the predicted probability of a QC rule rejection when the measurement procedure is in its stable in-control state is low, such as 0.01. A common application of the range rule is denoted R_{4s} and rejects if the range of the z-scores or SDIs of the QC results from the current QC event exceeds 4. The range rule is designed to detect increases in imprecision and has little ability to detect shifts.

Individual QC rules vary in their ability to detect different types and sizes of out-of-control conditions. QC rules based on QC results from a single QC event, such as the 1_{3s} rule, are best at detecting large shifts quickly and also have the ability to detect increases in imprecision. Counting rules, such as the 10_{1s} rule, and the moving mean, EWMA, and CUSUM rules, are designed to detect smaller shifts and are particularly good at detecting trends.

Individual QC rules are often combined into a QC multirule. A QC multirule rejects if any of the individual QC rules reject. For example, a $1_{3s}/2_{2s}/4_{1s}$ multirule evaluates all three individual counting rules and rejects if any of the individual rules reject. Examples of QC multirules are:

- $1_{3s}/2_{2s}/R_{4s}$ evaluating two QC concentrations
- $1_{3s}/2_{2s}/R_{4s}/8_{1s}$ evaluating two QC concentrations
- 1_{3s}/2 of $3_{2s}/R_{4s}$ evaluating three QC concentrations
- 1_{3s}/2 of $3_{2s}/R_{4s}/6_{1s}$ evaluating three QC concentrations

With computerized analytical and information systems, it is now practical to use more complex statistical rules, such as multirules, rules that entail transforming QC results into z-scores or SDIs (eg, the mean rule), or rules that use current and previous QC results (eg, the moving mean, EWMA, and CUSUM rules).

Selection of QC rules to use for evaluation of the QC results for a given measurement procedure is based on the considerations described in Chapters 4 and 5. Power function graphs or computer simulation can be used to predict the performance of QC rules under the assumptions used for the calculations. The SD used for the calculations is critical and needs to be a good estimate of the overall long-term variability of the measurement procedure when it is operating in a stable condition. Most software programs assume a gaussian (normal) distribution of results that may not be appropriate for some types of long-term components of variability such as periodic recalibration or maintenance procedures.

Empirical evaluation of QC rules performance can also be done by obtaining a large series of QC results from a measurement procedure that has been operating in a stable condition over a sufficiently long time interval to include all major sources of variability in the data.[5] Each candidate QC rule is applied to the data to determine the false-positive rate and the frequency of detecting (which represents the probability of detecting) bias error conditions of a specified magnitude introduced into the data. The empirical approach is similar to computer simulation but offers an opportunity to assess QC rules performance based on actual QC results that reflect all of the types of variability observed over time.

The final choice of a set of QC rules is made to have as low a false rejection rate as possible while having the ability to detect error conditions large enough to cause a hazardous condition that may affect patient care decisions, ideally before erroneous patient results are reported and acted on. Laboratories should use a set of QC rules that are sensitive to different types of out-of-control conditions to increase the probability of detecting different error conditions.

For measurement procedures with very good analytical performance compared to the medical needs (eg, a 5 to 6 Sigma level), a less stringent QC rule, such as 1_{3s} or 1_{4s}, will likely have a high probability of detecting an error that may cause a hazardous condition while providing a very low false rejection rate.

For measurement procedures with marginal analytical performance compared to the medical needs (eg, a 2 to 3 Sigma level), a combination of QC rules will likely be needed. Several QC concentrations may be used to improve the probability of detecting an error because a small magnitude error may represent a hazardous condition. The laboratory may have to accept a higher false rejection rate for a poorer performing measurement procedure in order to increase the probability of detecting an error condition. In addition, the probability of detecting an error condition may be less than ideal, which may increase the time to detect an error condition and/or the number of patient results reported before an error condition is identified. In these situations, the laboratory may consider increasing the frequency of performing QC events in order to reduce the number of patient results reported before an error condition is identified.

In addition to various QC rules based on multiples of SD, using a trend detection rule such as CUSUM, EWMA, or a similar rule is recommended. Trend detection rules are particularly helpful for continuous measurement situations and for measurement procedures with poorer analytical performance relative to the medical needs. The threshold for an alert using trend rules can be set such that a developing analytical error condition can be detected before it is large enough to cause a hazardous condition for patients. Used in this manner, a trend rule could be used as a warning rule that may not warrant immediately discontinuing use of a measurement procedure. Alternatively, the threshold for an alert from a trend rule can be set to identify a change in performance that would necessitate immediate discontinuation of a measurement procedure.

5.5.2 Quality Control Schedules

This guideline assumes that a measurement procedure was selected by a laboratory to produce results that are fit for their intended use in diagnosis, treatment, or monitoring of one or more disease conditions. Consequently, QC samples should be measured as appropriate to allow the laboratory to have confidence that the measurement procedure is producing results for patient specimens that are consistent with the performance expectations of the procedure.

The laboratory should consider the following situations in determining when QC samples should be measured.

5.5.2.1 Batch Quality Control

Batch QC refers to the condition in which a group of patient specimens is measured by a procedure that is characterized by a defined start and stop time with all measurements occurring for all specimens during that time interval. One example is a microplate format that can accommodate a predefined number of samples (typically including patient specimens, calibrator, and control samples) that are analyzed as a unit. The samples and reagents may be pipetted individually and sequentially or in parallel using multichannel fluid handling devices that may influence the location of QC samples. Another example is a measurement procedure in which a defined number of samples are measured sequentially, typically with calibrators and controls included in the sequence. It is important to note that the time interval that is considered a batch may be short or may extend for many hours depending on the stability of the measurement procedure and/or the total number of samples to be measured.

In batch measurement mode, QC samples should be included such that the results for QC samples can be used to verify that the measurement procedure remained stable during the interval of the measurements and thus the results for the patient specimens are likely to be correct. The number and placement of QC samples is determined by considering the analytical stability of the measuring system over the interval of time needed to complete the batch. For a microplate format, including a minimum of two or three QC samples on the plate is recommended. For a sequential series of measurements, including QC samples at the beginning and end of the series and considering QC samples at other positions according to the stability of the measurement procedure is recommended.

5.5.2.2 Continuous Quality Control

A continuous measurement process occurs without defining a specific time interval for the measurements and typically continues indefinitely until an event such as reagent replenishment, recalibration, or maintenance occurs. In continuous mode, QC samples are measured periodically along with patient specimens. QC results from the current QC event are interpreted to reflect the current condition of the measurement procedure. If the current QC sample results are acceptable, it is assumed that the measurement procedure has remained stable since the last acceptable QC event, and thus, the results for patient specimens measured during that interval are likely to be acceptable. This type of QC schedule can be called "bracketed QC" because the results at the beginning and end of a "bracket" are used to verify that patient results measured within the "bracket" are acceptable.

The frequency of QC sample measurements in a continuous mode, or the interval of a bracket, is influenced by several considerations.[55,56] Subchapters 5.5.2.3 through 5.5.2.6 focus on the common considerations. These considerations are not mutually exclusive and influence each other. The final determination of frequency to perform measurements on QC samples is based on integrating the various considerations to allow verification that the results for patient specimens are likely to be correct and thus suitable for their intended use in patient care.

5.5.2.3 Critical Control Point Quality Control

There are scheduled events that could alter the performance of a measurement procedure. These are referred to as critical control points. Examples include:

- Calibration
- Maintenance
- A new container of the same lot of a reagent
- Reagent lot change
- Calibrator lot change

The laboratory should consider if such events could sufficiently alter the measurement conditions to cause results for patient specimens to be unacceptable for their intended use in clinical care.

When operating in batch mode and no such event occurs during a batch, no additional QC measurements are needed. In this situation, the QC measurements associated with a batch are adequate to demonstrate that the measurement procedure most likely performed correctly and that the results for patient specimens are acceptable.

When operating in a continuous mode and a critical control point event occurs, it is necessary to verify the performance of the measurement procedure both before and after the event. In continuous mode, QC samples are measured periodically along with patient specimens. The results for the current QC samples reflect the current condition of the measurement procedure and are thus used to verify the likelihood that the results for patient specimens have remained acceptable since the last time QC measurements were made. Consequently, if a critical control point event is scheduled to occur, it is necessary to verify the

condition of the measurement procedure before the conditions are altered by the event. Otherwise, there is no valid information to conclude that there had been no change in the acceptability of the patient results reported since the last QC measurements. It is also necessary to perform QC sample measurements after the critical control point event to verify the event was successful and did not unintentionally alter the measurement conditions causing the results for patient specimens to no longer be acceptable.

If a critical control point event occurs that is not scheduled and interrupts the continuous mode of a measurement procedure, then the laboratory may not have QC information to evaluate whether the patient results are likely to be correct since the last QC measurements were made. In this situation, the laboratory needs to consider the likelihood that patient results from before the event could be erroneous, and determine whether to remeasure the patient specimens to confirm the acceptability of values that may have already been reported.

5.5.2.4 Stability of the Measurement Procedure

The stability of a measurement procedure can be demonstrated by minimal drift and no or insignificant shifts in a Levey-Jennings graph (see Appendix A) over an extended time interval. The magnitude of drift and/or shifts that is considered insignificant depends on the clinical use of a result (see Subchapters 3.2 and 3.3).

The less stable the measurement procedure, the more frequently QC samples should be examined. QC results are needed at a frequency that confirms that a measurement procedure has remained stable and that the results for patient specimens are likely to be acceptable for their intended clinical use. QC results are also needed at a frequency that alerts the technologist that an error condition has occurred and corrective action, including repeating patient specimens, is needed. Ideally the error condition is detected before releasing the patient results, but should definitely be detected before continuing the measurement process. Any patient results already released that are determined to be erroneous need to have corrected reports issued.

5.5.2.5 Number of Patient Results Expected to Be Reported Between Quality Control Events

The frequency of QC events may be determined by considering the number of patient results reported between QC events. If an error condition is detected by the next QC event, corrective action is needed. Corrective action includes determining which patient results are erroneous and issuing corrected reports. The laboratory should consider the cost of increased frequency of QC events vs the time and cost to repeat the patient specimen measurements and to issue corrected reports, as well as the risk of harm.

5.5.2.6 Risk of Harm to a Patient if an Erroneous Result Is Reported

The frequency of measuring QC samples may be determined primarily by the risk that an erroneous result could cause harm to a patient before the error condition would be identified by the next scheduled QC event and the patient specimen result could be corrected.

It is important to note that a relationship exists between the stability of a measurement procedure and the likelihood that an erroneous result may occur. The laboratory should consider the types of malfunctions that may occur and the frequency at which they could occur when assessing the risk of harm to a patient should an erroneous result be reported and acted on. These considerations are helpful in assessing the frequency for bracketed QC. Control procedures that may be built into a measurement procedure should also be considered when assessing the risk for harm (see CLSI document EP23[2]).

5.6 Design a Quality Control Strategy for Multiple Instruments

Most of the laboratory medicine literature discussing statistical QC principles and practices considers the problem of monitoring the performance of a single measurement procedure using a single instrument. In many laboratories there may be multiple instruments of the same type performing the same menu of measurement procedures.

The problem of developing appropriate QC strategies for multiple instruments that measure the same measurands is one of the important challenges in the modern laboratory. Some of the additional factors that should be considered when designing a QC strategy for multiple instruments are:

- Differences in estimates of stable baseline analytical performance between multiple instruments

- The importance of an instrument's change in performance relative to its own baseline as well as to the stable baselines of the other instruments

- Whether to use the same QC target and SD for all instruments measuring the same measurand or use individual QC targets and SDs for each instrument

Because there is very limited literature discussing multiple instrument QC strategy design, there are no consensus recommendations for this situation. This is an area that would benefit from additional research (see CLSI document EP05[24]).

Chapter 6: Recovering From an Out-of-Control Condition

This chapter includes:

- Responding to out-of-control events and conditions
- Identifying and correcting any erroneous patient results

6.1 Responding to an Out-of-Control Quality Control Event

When a QC rule evaluation suggests the measurement process is out of control, the out-of-control results can be verified by repeating the measurements using fresh QC material in order to rule out any issues that could be caused by compromised QC material (eg, evaporation, unsuitable storage conditions, or sampling the wrong concentration of QC material [see Appendix B]). When fresh QC material reproduces the out-of-control condition, it should be handled as an authentic failure reflecting an analytical measurement issue. When fresh QC material does not reproduce the out-of-control condition, the laboratory should review recent QC values to determine if there may be a trend toward an out-of-control condition that needs to be considered. For example, if the results for a repeated fresh QC result and for the past several QC measurements (or for several attempts to repeat a fresh control) are very close to an out-of-control decision value, it is more likely that an out-of-control condition exists and less likely that the measurement procedure is in a stable condition. Repeating measurement of QC material should be used only to rule out obvious problems with the QC material itself. Continuing to repeat QC measurements with the intention of obtaining in-control results is an unacceptable practice.

6.2 Responding to an Out-of-Control Condition

When an out-of-control condition has been detected, the first step is to contain the error condition by immediately discontinuing patient testing and/or patient result reporting. In automated continuous testing scenarios, discontinuing testing can be achieved using middleware or laboratory information system functionality or by taking the analyzer/measurement procedure off-line and out of production. In laboratories using autoverification for reporting patient results, autoverification should be stopped as soon as an out-of-control QC event has occurred.

Intervention is needed to correct out-of-control conditions.[57] Examples of common corrective actions for out-of-control conditions include calibration, replacement of reagent containers, or replacing electrodes. Less common occurrences may call for more in-depth investigation to determine the root cause. Details surrounding the investigation and troubleshooting should be documented. After identification and correction of the root cause, QC should be measured to verify that stable operation has resumed.

6.3 Identifying and Correcting Reported Erroneous Patient Results

The medically significant magnitude of change that necessitates a corrected report should be defined for each measurand reported by the laboratory.

When more than one testing platform is available and in control, retesting patient specimens can be done in parallel to troubleshooting. The general approach is to repeat patient specimen testing using an in-control measurement procedure and compare the results to those originally reported. Differences between results that exceed the predetermined medically significant magnitude of change should be corrected, and corrected patient reports need to be issued. Repeat testing should begin from the point in time of the QC rule rejection that detected the error condition and continue back in time until the point in time that the error condition occurred. The length of time an error condition exists before it is detected is correlated with the magnitude of the error condition. A relatively small error condition may persist across multiple

QC events before it is finally detected. A large error condition is more likely to be detected at the first QC event after the error condition occurs.

It is not always possible to identify the exact point in time when the error occurred, but various strategies are used to identify which patient specimens need to be retested. One option is for the laboratory to retest all patient specimens measured since the last in-control QC event. This approach works well for batch testing, bracketed QC, or continuous testing scenarios for which scheduled QC testing intervals are relatively short or the number of patient specimens measured between QC events is small.

A second option is to retest patient specimens until the approximate point in time when the error condition began. This approach is accomplished by retesting batches of patient specimens or retesting patients at defined intervals. For example, the laboratory may retest patient specimens in batches of 10 going back in time to the last in-control QC event. If any of the 10 patient results need correction, the laboratory continues retesting another batch of 10 specimens. The retesting process continues in batches of 10 until an entire batch is encountered that needs no corrected results. This point in time approximates when the error condition occurred. Retesting selected patient specimens at defined intervals since the last acceptable QC event can also be used to approximate when the error condition began. Once the point in time when the error occurred is identified, it is important that all patient specimens tested during the out-of-control condition be retested to assess if the error was large enough to affect patient care.

Repeat testing should include patient specimens with measurand concentrations near the concentration at which the out-of-control error condition occurred. For example, if a laboratory is using a batch testing approach to repeat testing, and the first batch of patient specimens retested does not contain results near the concentration at which the out-of-control condition was detected, then repeat testing should continue targeting patient specimens with results near the concentration of the out-of-control QC material. Similarly, when retesting selected patient specimens at defined intervals, the concentrations should include those near the concentration at which the out-of-control result occurred in order to be confident in correctly identifying the time the out-of-control condition started.

If patient specimens are not available for repeat testing or the measurand is labile and cannot be retested, the laboratory should issue a corrected patient report indicating that the result is not valid. This information should be available in the patient’s medical record.

Chapter 7: Ongoing Assessment of Quality Control Programs

This chapter includes:

- Points to consider for the ongoing assessment of internal and interlaboratory QC programs

7.1 Assessment of the Internal Quality Control Program

In order to maximize error detection and minimize false rejection, ongoing assessment of a laboratory's QC program is necessary to ensure the QC program is serving its intended purpose.

Examples of ongoing assessment and review of QC data include:

- Periodically reviewing the mean, SD, and CV to ensure an appropriate target value and SD are used, and to identify changes in method performance that may need corrective action

- Investigating measurement procedures with frequent QC failures to determine the root cause of the failures and to identify corrective action

- Monitoring the rate of QC rule rejections and the number of patient specimens needing retesting due to QC rule rejections compared to the number of patient results requiring correction

- Reviewing the analytical errors that were not detected using statistical QC to determine whether the QC strategy can be modified to detect the error, if it should occur again

These data can be used collectively to optimize the QC frequency and the selection of statistical rules.

It is important to reassess the laboratory's QC plan when changes occur in the laboratory. For example, new instrumentation or an increase or decrease in measurement procedure volume may warrant changes in the QC program.

7.2 Using Interlaboratory Quality Control to Assess a Quality Control Program

An interlaboratory QC program is a means for statistically evaluating the performance of a measurement procedure by comparing results for QC materials to the results for the same (ie, identical lot number) QC materials measured by like (or substantially like) measurement procedures in other laboratories.

Some advantages for participating in an interlaboratory QC program are:

- Verifying that a laboratory is producing QC results that are consistent with other laboratories using the same measurement procedure, and thus demonstrating that the laboratory is using the measurement procedure correctly

- Detecting and identifying bias in a measurement procedure
  - Bias can be caused by events such as reagent or calibrator lot changes or reformulations, changes in calibration traceability to reference systems, or instrument software changes. Comparison of an individual laboratory's QC result to a peer group mean value can identify a trend or shift, or ascertain if other laboratories are experiencing the same changes (see Subchapter 3.3.1).

- Supplementing PT/EQA programs
  - PT/EQA programs verify performance at a point in time. Acceptable performance on the day of PT/EQA testing does not guarantee testing reliability every day because errors in a measurement procedure can occur at any time. In addition, the interlaboratory QC data can be used to investigate a failed or questionable PT/EQA result.

Chapter 8: Worked Examples

This chapter includes:

- Complete examples for setting quality requirements and defining QC strategies for two measurement procedures:
  - Thyroid-stimulating hormone (TSH)
  - Calcium

This chapter works through two examples that illustrate the entire process from setting the quality requirement to defining the QC strategy. The measurands selected for the examples are TSH and calcium, in order to show different possible choices and outcomes for the QC strategy. These examples illustrate setting the quality requirement, assigning a target value and SD for a new lot of QC material, and selecting appropriate rules for interpreting QC results.

NOTE: These are representative examples and are not meant to be recommendations regarding QC for these particular measurands. Each laboratory should evaluate their individual needs and/or requirements.

8.1 Define the Quality Requirement

The quality requirement is set using a clinically based goal deemed appropriate by the laboratory.

TSH: After review of possible sources for a quality goal, consideration of biological variability is deemed appropriate for TSH. As noted in published tables for biological variability, the TEa goal is ± 23.7% (see Tables 3 and 4).

Calcium: Review of biological variability data for calcium shows a TEa goal of 2.6%. Review of performance data for the calcium measurement procedure indicates that this goal is not realistically achievable with the current technology. Alternate goals are evaluated based on consultation with clinical care providers, and a TEa goal of ± 6% is chosen.

8.2 Select Quality Control Materials

For both TSH and calcium, commercially available QC materials are selected due to appropriate concentrations, product stability, and shelf life. Three concentrations are available for TSH because both very low and high values are medically meaningful. Two concentrations are available for calcium because the measuring interval is relatively small and performance is similar at all concentrations.

8.3 Determine Target Values and Standard Deviations

For both measurands, each QC material is measured once a day for 10 days, and the average result is used as the initial estimate for the target value. The SDs for the new lots of QC materials are based on the SDs calculated for the previous lots using data accumulated over the period of use as described in Subchapter 5.3.1.

Table 3. Target Values, SDs, and CVs for the QC Materials in This Example

Measurand	Level	Target		SD		CV
TSH		mIU/L		mIU/L		%
	1	0.12		0.0053		4.41
	2	0.85		0.022		2.58
	3	5.20		0.130		2.53
Calcium		mmol/L	mg/dL	mmol/L	mg/dL	
	1	2.55	10.20	0.036	0.143	1.40
	2	3.24	12.96	0.048	0.193	1.49

Abbreviations: CV, coefficient of variation; QC, quality control; SD, standard deviation; TSH, thyroid-stimulating hormone.

8.4 Select Quality Control Strategy

A QC strategy is based on the desired high probability of detecting a significant change in measurement procedure performance and the desired low probability of false rejections. A comparison of measurement procedure performance to the quality requirement is made to select appropriate QC rules. Sigma metric values (see Subchapter 3.3.3) are estimated to characterize the measurement procedure performance. The Sigma metric at a specific concentration *x* can be estimated as:

$$\text{Sigma}(x) = \frac{\text{TEa}(x) - |\text{Bias}(x)|}{\text{SD}(x)} \quad (4)$$

Because bias data vs a reference measurement procedure are difficult to obtain, bias is assumed to be zero, as discussed in Subchapter 3.3.1.

Table 4. Sigma Metrics for the Measurement Procedures in This Example

Measurand	Concentration		TEa		SD		Sigma Metric
TSH	mIU/L		mIU/L		mIU/L		
	0.12		0.028		0.0053		5.3
	0.85		0.20		0.022		9.1
	5.20		1.23		0.131		9.4
Calcium	mmol/L	mg/dL	mmol/L	mg/dL	mmol/L	mg/dL	
	2.55	10.20	0.153	0.612	0.036	0.143	4.3
	3.24	12.96	0.194	0.776	0.048	0.193	4.0

Abbreviations: SD, standard deviation; TEa, allowable total error; TSH, thyroid-stimulating hormone.

The estimated Sigma metrics for TSH suggest that substantial changes in measurement procedure performance can occur before a change affects medical decisions. The Sigma value for the lower concentration is not as favorable as for the higher concentration; therefore, the QC strategy is designed around the lower concentration performance. Thus, as discussed in Subchapter 4.1, the QC rules can be chosen to minimize false rejections, while still having a good probability to detect a change in the measurement procedure's performance before the TEa is exceeded. A single QC rule, such as 1_{3s}, is suitable to detect a clinically meaningful change in performance.

The estimated Sigma metrics for calcium suggest that small changes in measurement procedure performance may affect medical decisions. Identifying small changes in performance entails more complex QC rules involving a multirule approach designed to increase the probability of detecting a change in measurement procedure performance, while keeping false rejections to a tolerable frequency. A candidate strategy is using 1_{3s}, 2_{2s}, and R_{4s} rules together with two QC concentrations at every QC event. Adding a CUSUM or EWMA rule would also be useful to identify trends before a significant error condition might occur.

The frequency of measuring QC samples is based on two concepts. First, critical control point QC should be considered, which is testing QC samples before and after every scheduled event that may affect measurement procedure performance, such as calibration, maintenance, or starting a new reagent lot. QC samples should always be measured before and after these events. Second, QC samples should be measured at a scheduled frequency during routine operations to detect occurrence of a measurement procedure failure. That frequency is determined to control risk of reporting undetected erroneous results that have a risk to cause harm to a patient. One factor in limiting risk is the number of patient specimens analyzed in a time interval. A larger number of patient specimens measured between QC events increases the risk for erroneous results being reported and used for a medical decision before a QC rule evaluation can detect an error.

For TSH, the Sigma metric of approximately 5 to 9 suggests that substantial changes in measurement procedure performance could occur before the magnitude of an erroneous result would alter medical decisions. Additionally, a single erroneous TSH result carries a low risk of an immediate medical treatment or nontreatment decision that might cause harm to a patient. In this example, approximately 200 TSH measurements are made during one eight-hour interval per day. The laboratory chooses to measure QC samples at the beginning and end of the eight-hour interval. In the event a measurement procedure problem is identified by a QC event at the end of the eight-hour interval, the 200 patient results can be repeated and corrected reports issued. This timeliness of corrected results' availability is considered acceptable by the laboratory director because an erroneous result could most likely be corrected before a patient intervention would have occurred. This QC strategy, including number of controls, rule selection, and frequency of measurement, is determined by the laboratory director to meet the needs of the patients served.

For calcium, the Sigma metric of approximately 4 suggests that small changes in measurement procedure performance may alter medical decisions. A single erroneous calcium result could cause an immediate clinical intervention decision with potentially serious harm to a patient. Therefore, the QC strategy for calcium should focus on quickly detecting small magnitude changes in measurement procedure performance to minimize the risk of harm to patients. More sensitive QC rules, testing multiple QC samples at each QC event, and more frequent QC events all contribute to reducing patient risk. In this example, approximately 500 calcium measurements are made during a 24-hour interval, seven days per week. The laboratory chooses to measure QC samples at intervals of six hours throughout the 24-hour interval. In the event a measurement procedure problem is identified by a QC event, patient results from the previous six hours (approximately 125 results) are repeated and corrected reports issued, if needed. This timeliness of corrected results availability is considered acceptable by the laboratory director. This QC strategy, including number of controls, rule selection, and frequency of measurement, is determined by the laboratory director to meet the needs of the patients served by the laboratory.

Chapter 9: Conclusion

Key messages of C24 include:

- The main goal of laboratory QC is to reduce the risk of harm to a patient associated with an erroneous result.

- QC strategies are an important component of assessing measurement procedure performance.

- In choosing a QC strategy, the laboratory should give serious consideration to the choice of QC materials, the QC rules evaluated, and the frequency at which QC events are scheduled.

- There is not one QC strategy that is best for all measurement procedures. The preferred QC strategy for a measurement procedure takes into account:
  - The quality required for patient results
  - A measurement procedure's performance capability relative to the quality required
  - The likelihood and severity of harm to the patient if an erroneous result is acted on inappropriately
  - The stability of a measurement procedure

- The laboratory should identify and correct reported erroneous patient results after an out-of-control condition in a measurement procedure has been detected.

Finally, although significant advances in QC thinking have occurred, there are still important areas that could benefit from additional development, such as QC strategy design and implementation for laboratories with multiple instruments of the same type performing the same measurement procedures.

Chapter 10: Supplemental Information

This chapter includes:

- References
- Appendixes
- The Quality Management System Approach
- Related CLSI Reference Materials

References

1 ISO. *Medical laboratories -- Requirements for quality and competence.* ISO 15189. Geneva, Switzerland: International Organization for Standardization; 2012.

2 CLSI. *Laboratory Quality Control Based on Risk Management*; *Approved Guideline.* CLSI document EP23-A™. Wayne, PA: Clinical and Laboratory Standards Institute; 2011.

3 Westgard JO. *Six Sigma Quality Design & Control: Desirable Precision and Requisite QC for Laboratory Measurement Processes.* 2nd ed. Madison, WI: Westgard QC, Inc.; 2006.

4 Burnett D, Ceriotti F, Cooper G, Parvin C, Plebani M, Westgard J. Collective opinion paper on findings of the 2009 convocation of experts on quality control. *Clin Chem Lab Med.* 2010;48(1):41-52.

5 Miller WG. Quality control. In: McPherson RA, Pincus, MR. *Henry's Clinical Diagnosis and Management by Laboratory Methods*. 22nd ed. Philadelphia, PA: Elsevier Saunders; 2011:119-134.

6 Cooper G, DeJonge N, Ehrmeyer S, et al. Collective opinion paper on findings of the 2010 convocation of experts on laboratory quality. *Clin Chem Lab Med.* 2011;49(5):793-802.

7 Klee GG, Westgard JO. Quality management. In: Burtis CA, Ashwood ER, Bruns DE. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics.* 5th ed. Philadelphia, PA: Elsevier- Saunders; 2012:163-203.

8 Adams O, Cooper G, Fraser C, et al. Collective opinion paper on findings of the 2011 convocation of experts on laboratory quality. *Clin Chem Lab Med.* 2012;50(9):1547-1558.

9 Westgard JO, Westgard S, eds. *Quality Control in the Age of Risk Management. Clinics in Laboratory Medicine.* 2013;33(1):1-206.

10 Miller JM, Astles JR, Baszler T, et al.; Biosafety Blue Ribbon Panel, Centers for Disease Control and Prevention (CDC). Guidelines for safe work practices in human and animal medical diagnostic laboratories. *MMWR Surveill Summ.* 2012;61 Suppl:1-102.

11 CLSI. *Protection of Laboratory Workers From Occupationally Acquired Infections; Approved Guideline—Fourth Edition.* CLSI document M29-A4. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

12 Bureau International des Poids et Mesures (BIPM). *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms* (VIM, 3rd edition, JCGM 200:2012). http://www.bipm.org/en/publications/guides/vim.html. Accessed June 28, 2016.

13 ISO. *In vitro diagnostic medical devices – Information supplied by the manufacturer (labelling) – Part 1: Terms, definitions and general requirements.* ISO 18113-1. Geneva, Switzerland: International Organization for Standardization; 2009.

14 ISO. *Statistics – Vocabulary and symbols – Part 2: Applied statistics.* ISO 3534-2. Geneva, Switzerland: International Organization for Standardization; 2006.

15 ISO. *Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability.* ISO 3534-1. Geneva, Switzerland: International Organization for Standardization; 2006.

16 IEC. *International Electrotechnical Vocabulary – Electrical and electronic measurements and measuring instruments*. IEC 60050-300. Geneva, Switzerland: International Electrotechnical Commission; 2001.

17 ISO. *Quality management systems – Fundamentals and vocabulary.* ISO 9000. Geneva, Switzerland: International Organization for Standardization; 2015.

18 ISO. *Accuracy (trueness and precision) of measurement methods and results – Part I: General principles and definitions.* ISO 5725-1. Geneva: International Organization for Standardization; 1994.

19 ISO. *In vitro diagnostic medical devices – Measurement of quantities in samples of biological origin – Requirements for content and presentation of reference measurement procedures.* ISO 15193. Geneva, Switzerland: International Organization for Standardization; 2009.

20 Bureau International des Poids et Mesures (BIPM). *Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement* (GUM, 1st edition, JCGM 100:2008). http://www.bipm.org/en/publications/guides/gum.html. Accessed August 1, 2016.

21 CLSI. *Measurement Procedure Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Third Edition.* CLSI document EP09-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2013.

[22] Bureau International des Poids et Mesures. JCTLM-WG2: Reference measurement laboratories. http://www.bipm.org/en/committees/cc/wg/jctlm-wg2.html. Accessed June 28, 2016.

[23] Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem.* 2011;57(12):1670-1680.

[24] CLSI. *Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline—Third Edition.* CLSI document EP05-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

[25] Westgard JO, Westgard SA. Total analytic error: from concept to application. *Clin Lab News.* 2013;9:8-10.

[26] Parvin CA. Comparing the power of quality-control rules to detect persistent systematic error. *Clin Chem.* 1992;38(3):358-363.

[27] Parvin CA. Comparing the power of quality-control rules to detect persistent increases in random error. *Clin Chem.* 1992;38(3):364-369.

[28] Westgard JO. Internal quality control: planning and implementation strategies. *Ann Clin Biochem.* 2003;40(Pt 6):593-611.

[29] Parvin CA, Gronowski AM. Effect of analytical run length on quality-control (QC) performance and the QC planning process. *Clin Chem.* 1997;43(11):2149-2154.

[30] Kallner A, McQueen M, Heuck C. The Stockholm Consensus Conference on quality specifications in laboratory medicine, 25-26 April 1999. *Scand J Clin Lab Invest.* 1999;59(7):475-476.

[31] Petersen PH, Fraser CG, Kallner A, Kenny D, eds. *Scand J Clin Lab Invest.* 1999;59(7, special issue):475-585.

[32] Sandberg S, Fraser CG, Horvath AR, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem Lab Med.* 2015;53(6):833-835.

[33] Plebani M, ed. 1st EFLM Strategic Conference / Defining analytical performance goals – 15 years after the Stockholm Conference. *Clin Chem Lab Med.* 2015;53(6, special issue):829-958.

[34] Horvath AR, Bossuyt PM, Sandberg S, et al.; Test Evaluation Working Group of the European Federation of Clinical Chemistry and Laboratory Medicine. Setting analytical performance specifications based on outcome studies–is it possible? *Clin Chem Lab Med.* 2015;53(6):841-848.

[35] Petersen PH. Performance criteria based on true and false classification and clinical outcomes: influence of analytical performance on diagnostic outcome using a single clinical component. *Clin Chem Lab Med.* 2015;53(6):849-855.

[36] Thue G, Sandberg S. Analytical performance specifications based on how clinicians use laboratory tests: experiences from a post-analytical external quality assessment programme. *Clin Chem Lab Med.* 2015;53(6):857-862.

[37] Fraser CG. *Biological Variation: From Principles to Practice.* Washington, DC: AACC Press; 2001.

[38] Ricós C, Álvarez V, Perich C, et al. Rationale for using data on biological variation. *Clin Chem Lab Med.* 2015;53(6):863-870.

[39] Perich C, Minchinela J, Ricós C, et al. Biological variation database: structure and criteria used for generation and update. *Clin Chem Lab Med.* 2015;53(2):299-305.

[40] Oosterhuis WP. Gross overestimation of total allowable error based on biological variation. *Clin Chem.* 2011;57(9):1334-1336.

[41] Røraas T, Petersen PH, Sandberg S. Confidence intervals and power calculations for within-person biological variation: effect of analytical imprecision, number of replicates, number of samples, and number of individuals. *Clin Chem.* 2012;58(9):1306-1313.

[42] Aarsand AK, Rørass T, Sandberg S. Biological variation – reliable data is essential. *Clin Chem Lab Med.* 2015;53(2):153-154.

[43] Panteghini M, Sandberg S. Defining analytical performance specifications 15 years after the Stockholm conference. *Clin Chem Lab Med.* 2015;53(6):829-832.

[44] Carobene A. Reliability of biological variation data available in an online database: need for improvement. *Clin Chem Lab Med.* 2015;53(6):871-877.

[45] CLSI. *Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision).* CLSI document EP31-A-IR. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.

[46] CLSI. *User Evaluation of Between-Reagent Lot Variation; Approved Guideline.* CLSI document EP26-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2013.

[47] CLSI. *Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline.* CLSI document EP06-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2003.

[48] CLSI. *User Verification of Precision and Estimation of Bias; Approved Guideline—Third Edition.* CLSI document EP15-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

[49] Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. *Clin Chem.* 2011;57(1):76-83.

[50] Parvin CA. New insight into the comparative power of quality-control rules that use control observations within a single analytical run. *Clin Chem.* 1993;39(3):440-447.

[51] Parvin CA, Kuchipudi L, Yundt-Pacheco JC. Should I repeat my 1:2s QC rejection? *Clin Chem.* 2012;58(5):925-929.

[52] Neubauer AS. The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation. *Clin Chem.* 1997;43(4):594-601.

[53] Crowder SV. Design of exponentially weighted moving average schemes. *J Qual Technol.* 1989;21(3):155-162.

[54] Gan FF. An optimal design of CUSUM quality control charts. *J Qual Technol.* 1991;23(4):279-286.

[55] Parvin CA, Robbins S 3rd. Evaluation of the performance of randomized versus fixed time schedules for quality control procedures. *Clin Chem.* 2007;53(4):575-580.

[56] Parvin CA. Assessing the impact of the frequency of quality control testing on the quality of reported patient results. *Clin Chem.* 2008;54(12):2049-2054.

[57] Valenstein PN, Alpern GA, Keren DF. Responding to large-scale testing errors. *Am J Clin Pathol.* 2010;133(3):440-446.

Appendix A. Levey-Jennings Chart

Abbreviations for Appendix A

CV	coefficient of variation
QC	quality control
SD	standard deviation
SDI	standard deviation interval

For a quantitative measurement procedure, imprecision refers to the random variability (dispersion or error) in repeated measurements of a sample under fixed conditions. This dispersion is most usefully quantified in terms of SDs or measures derived therefrom, such as CVs.

A Levey-Jennings chart puts this variability on display by plotting the measurements in the time sequence in which they were generated. The name is derived from a journal article by Levey and Jennings credited with introducing medical laboratories to the Shewhart control charting techniques widely used to monitor industrial manufacturing processes (see CLSI document EP05[1]).[2-4]

As statistical QC practices have evolved within the clinical chemistry community, a prototypical Levey-Jennings chart (see Figure A1) is commonly understood as a plot of individual measurements on a quantitative scale (vertical axis) representing concentration (more generally, measurand value) against time on an ordinal scale (horizontal axis). Characteristically, the plot includes labeled tick marks and/or horizontal lines representing a second quantitative scale translating concentration values into deviations from a mean in SD units. This is referred to as a standard deviation interval (SDI) or z-score scale; it has also been called a Levey-Jennings scale (see CLSI document EP05[1]).

When a measurement procedure is operating normally, ie, exhibiting stable, in-control performance, and repeated measurements for a sample of suitable composition are generated under a particular set of conditions, the dispersion on display in a Levey-Jennings chart corresponds to the measurement procedure's inherent imprecision for samples with essentially the same composition, measurand value, and sources of variation at work under that set of conditions. Moreover, an SD calculated from those measurements constitutes an estimate of that characteristic.

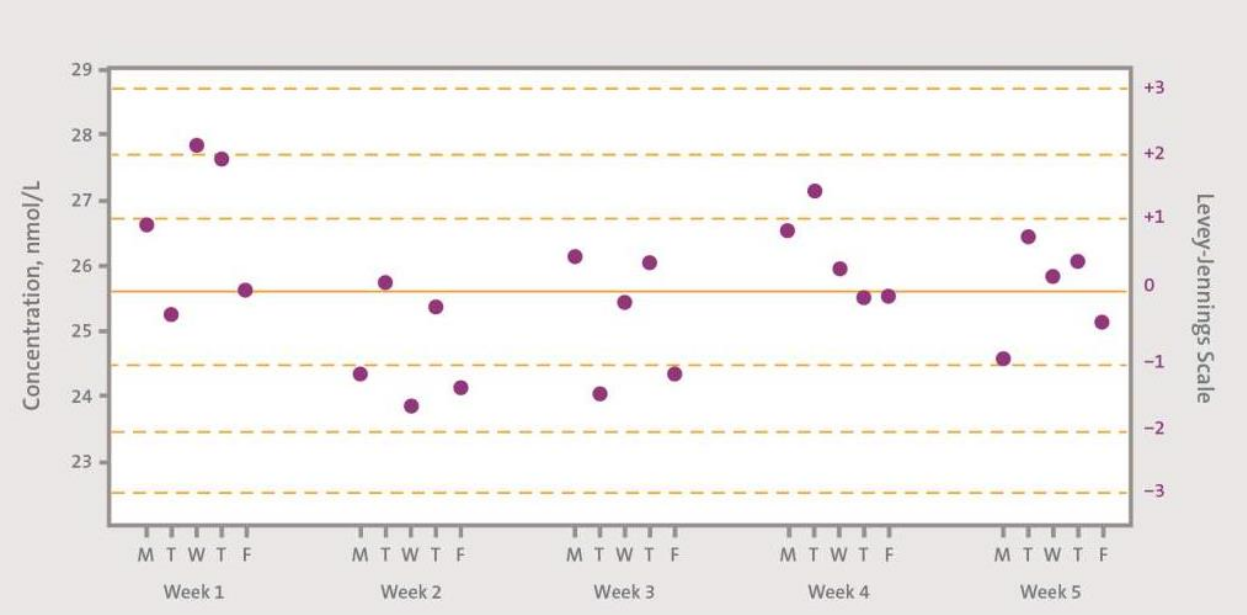
Applications. The importance of Levey-Jennings charts lies in their use for visually screening datasets intended as the basis for initial or updated mean and SD assignments, as discussed in Subchapter 5.3.1 of this guideline, and for surveying historical QC data. Statistical QC software applications commonly display such charts, which can be especially informative when the results are judiciously annotated and suitably aggregated across time, measurand values, QC materials, and relevant events.

Variations. The y-axis showing the values for QC results can be expressed in different ways. The most common is to show the mean and 1, 2, and 3 SD lines, in which the mean represents the target value for the QC sample and the SD represents the SD consistent with stable, in-control performance of the measurement procedure. The mean and SD may also be calculated from the data on display or an initial segment thereof, or the scale may be omitted. The SD scale is sometimes labeled with, or replaced by, percentiles for a gaussian distribution. For example, + 3 SDs corresponds roughly to the 99th percentile. The x-axis represents time and can be shown with different time increments such as actual year, month, day, hour, or minute, as appropriate. Alternatively, the x-axis can show relative time increments between individual observations.

To accommodate multiple QC samples with different measurand values in a single figure, charts may be stacked with their time scales suitably aligned. They may also be squeezed into a single chart and aligned on the SDI scale with either multiple concentration scales or no concentration scale. These and many other variations on the basic Levey-Jennings format may prove useful in particular situations.

Appendix A. (Continued)

For Levey-Jennings plots depicting historical QC data, it is often helpful to relate the ordinal time scale to calendar dates and to identify events potentially relevant to making sense of the data stream, such as changes in reagent lots, calibrator lots, or control materials; major maintenance events; decisions to update mean or SD assignments; or modifications of the QC interval or QC rules. This guideline provides instructive examples of Levey-Jennings plots spanning months and years (see Figures A2 and A3).

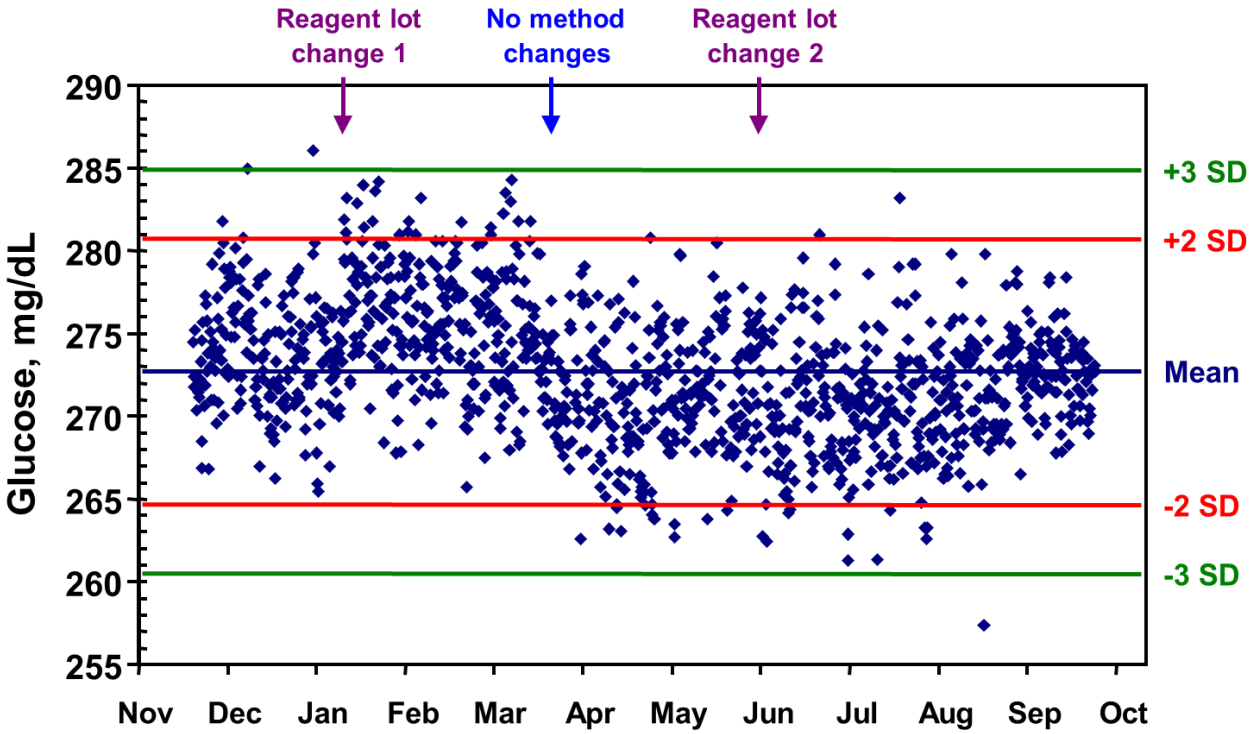


Abbreviations: M T W T F, Monday Tuesday Wednesday Thursday Friday.

Figure A1. Levey-Jennings Plot Example 1, Ferritin (Simulated Data)

Figure A1 depicts measurement results generated over five weeks to establish initial values for a QC material. Concentration is represented on the left vertical axis, and time points are on the horizontal axis. The horizontal lines, associated with the Levey-Jennings scale on the right vertical axis, indicate deviations from the mean in SD units based on statistics calculated for the results on display. The data points exhibit a pattern suggesting a source or sources of variability operating on a weekly basis; eg, weekly calibration and/or maintenance events, in addition to day-to-day (and within-day) sources (see CLSI document EP05[1]).

Appendix A. (Continued)

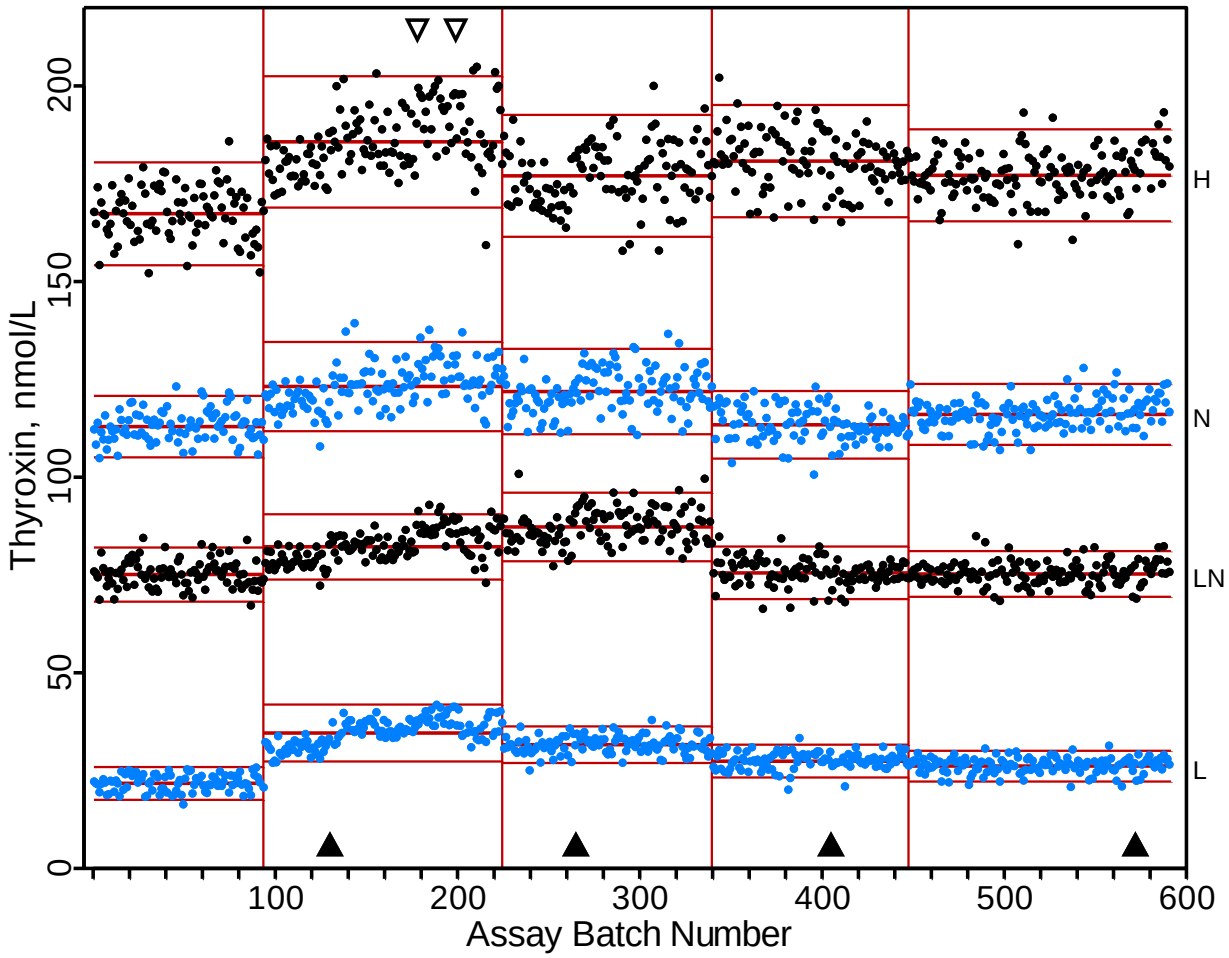


Abbreviation: SD, standard deviation.

Figure A2. Levey-Jennings Plot Example 2, QC Results. This figure was published in Miller WG. Quality control. In: McPherson RA, Pincus MR. *Henry's Clinical Diagnosis and Management by Laboratory Methods*. 22nd ed. Philadelphia, PA: Elsevier Saunders; 2011:119-134, and has been reprinted with permission.[5] Original source of figure: Reprinted with permission from Miller WG, Nichols JH. Quality control. In: Clarke W, ed. *Contemporary Practice in Clinical Chemistry*. 2nd ed. Washington, DC: AACC Press; 2011:57-71.[6]

Figure A2 shows a Levey-Jennings plot of QC results ($N = 1232$) for a single lot of QC material used over a 10-month period. The mean determined from the results for the first 49 days and the cumulative SD for the 10-month interval were used to label the y-axis. The data show that there were subintervals when the dispersion of results was smaller or larger and that small shifts in results occurred due to various unidentified influences on the measurement procedure. Noted on the figure is a small shift at the first reagent lot change, no influence of the second reagent lot change, and an unexplained small decrease in values between March and April.

Appendix A. (Continued)



Abbreviations: H, high; L, low; LN, low normal; N, normal.

Figure A3. Levey-Jennings Plot Example 3, Thyroxin (T4). Modified from Figure 1 in William A. Sadler, Murray H. Smith, Lynda M. Murray, and John G. Turner. A pragmatic approach to estimating total analytical error of immunoassays. Clinical Chemistry 1997; v. 43, p.608-614 (with permission from the American Association for Clinical Chemistry).[7]

Figure A3 shows results (means of duplicates) for a quad-level control generated in 591 consecutive in-control T4 radioimmunoassay batches over 29 months. Vertical lines indicate changes in the lots of QC materials. Horizontal lines represent means and 95% intervals calculated retrospectively for each lot of QC material. Closed arrows indicate calibrator lot turnovers. The two open arrows indicate statistically significant effects possibly associated with reagent lot changes.

References for Appendix A

1 CLSI. *Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline—Third Edition.* CLSI document EP05-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

2 Levey S, Jennings ER. The use of control charts in the clinical laboratory. *Am J Clin Pathol.* 1950;20(11):1059-1066.

3 Henry RJ, Segalove M. The running of standards in clinical chemistry and the use of the control chart. *J Clin Pathol.* 1952;5(4):305-311.

Appendix A. (Continued)

[4] Henry RJ. Use of the control chart in clinical chemistry. *Clin Chem.* 1959;5(4):309-319.

[5] Miller WG. Quality control. In: McPherson RA, Pincus MR. *Henry's Clinical Diagnosis and Management by Laboratory Methods.* 22nd ed. Philadelphia, PA: Elsevier Saunders; 2011:119-134.

[6] Miller WG, Nichols JH. Quality control. In: Clark W, ed. *Contemporary Practice in Clinical Chemistry.* 2nd ed. Washington, DC: AACC Press; 2011:57-71.

[7] Sadler WA, Smith MH, Murray LM, Turner JG. A pragmatic approach to estimating total analytical error of immunoassays. *Clin Chem.* 1997;43(4):608-614.

Appendix B. Medical Laboratory Quality Control Shift and Trend Troubleshooting Checklist

Measurand(s):		Analyzer(s):	
Technologist:		Technical/Quality Specialist:	
Medical Director:		Date:	

PURPOSE:
This checklist is used for investigating assignable cause(s) for shifts and/or trends in QC values in the medical laboratory. Relevant portions of the checklist may be used to aid laboratories' QC investigations.

Please indicate step completed by checking the "Done" box.

Table B1. Troubleshooting Checklist – Initial Information and Troubleshooting

Done	Task	Notes
☐	Review QC data (ie, Levey-Jennings charts) over the time interval during which the shift or trend occurred (eg, weeks, months). When did the shift or trend first occur? Did the shift or trend occur on one or more QC levels? Did the shift or trend occur for more than one measurand? Did the shift or trend occur on more than one analyzer?	
☐	Was a similar shift or trend observed for interinstrument patient comparison data?	
	QC Material Investigation	
☐	Is the QC material in use close to the open bottle expiration date? Was the bottle stored correctly between uses? Is the volume in the bottle low? Does the QC material have an abnormal appearance?	
☐	Is the shift or trend still observed when using fresh QC material (ie, trying a new bottle of QC material)?	

Appendix B. (Continued)

Table B1. (Continued)

Done	Task	Notes
☐	Was the QC from a new shipment of the same lot? Does the change in QC value(s) persist with a bottle from a different shipment (if available)?	
☐	Is the QC lot close to its expiration date?	
☐	Was a new QC material lot recently implemented? • Verify target values and SDs are correctly programmed in the instrument, computer system, or manual QC evaluation forms, as appropriate. • Review QC target value and SD assignment.	
☐	How does laboratory QC data compare to peer group data (if available)?	
	Reagent Investigation	
☐	Is the reagent close to its lot expiration date?	
☐	Was the reagent from a new shipment of the same lot number? Try reagent from a different shipment (if available).	
☐	Is the reagent a new lot recently implemented? Was the shift observed during the reagent validation?	
☐	When was the reagent prepared/loaded on the analyzer? • Inspect reagent for abnormalities. Does the change in QC value(s) persist after replacing the reagent with a new preparation/container?	
	Calibration Investigation	
☐	Was the measurement procedure recently calibrated? • Verify lot number and calibrator target value(s). • Review calibration data for error codes. • Did the calibrator have an abnormal appearance? Run calibrator as an unknown sample to compare results to target values. Do results compare?	

Appendix B. (Continued)

Table B1. (Continued)

Done	Task	Notes
☐	Does the change in QC value(s) persist after recalibration with a new bottle of calibrator? Was the calibrator from a new shipment of the same lot number? Does the change in QC value(s) persist after calibrating using a bottle from a different shipment (if available)?	
☐	Is the calibrator a new lot recently implemented? Does the shift/trend in QC value(s) persist after calibrating using a new lot of calibrator?	
	Analyzer Investigation	
☐	Was maintenance or service recently performed?	
☐	Have there been instrument alarms or analyzer malfunctions?	
☐	Are there any relevant product bulletins or recall notices?	
☐	Review reagent, calibrator, and QC product inserts. Are there any manufacturer recommendations that are not being followed?	
	Environmental Investigation	
☐	Were there any abnormal refrigerator/freezer alarms?	
☐	Were there any abnormal temperature or humidity alarms in the laboratory?	
☐	Are there any abnormal water quality readings? Send a water sample for testing.	
	Additional Investigation	
☐	Perform a sample comparison using one of the following: • Patient specimens tested before the QC shift • Patient specimens analyzed using another in-control instrument (if available) or at another laboratory using the same method • PT/EQA samples analyzed before QC shift (assuming the measurand is stable under the storage conditions) • Accuracy-based commutable reference or PT/EQA material with assigned values	

Appendix B. (Continued)

NOTES:_____

Table B2. Troubleshooting Checklist – Review Findings and Create a Corrective Action Plan

Done	Task	Notes
☐	Review findings.	
☐	Determine the next steps: Is there an assignable cause for the QC value(s) change? • If yes, has corrective action been taken? Is there evidence that patient results are not affected under the conditions of the QC value(s) change? • If patient results cannot be independently confirmed to be unaffected, is there evidence that all measurement procedure components are performing to specifications? Is additional validation or investigation needed? • Consult QC/analyzer manufacturer. • Adjust QC SD(s) if indicated. • Adjust QC target value(s) if indicated.	
☐	Implement changes.	
☐	Save documentation of investigation and action taken.	

NOTES:_____

Abbreviations: EQA, external quality assessment; PT, proficiency testing; QC, quality control; SD, standard deviation.

This page is intentionally left blank.

The Quality Management System Approach

Clinical and Laboratory Standards Institute (CLSI) subscribes to a quality management system (QMS) approach in the development of standards and guidelines, which facilitates project management; defines a document structure using a template; and provides a process to identify needed documents. The QMS approach applies a core set of "quality system essentials" (QSEs), basic to any organization, to all operations in any health care service's path of workflow (ie, operational aspects that define how a particular product or service is provided). The QSEs provide the framework for delivery of any type of product or service, serving as a manager's guide. The QSEs are as follows:

Organization	Personnel	Process Management	Nonconforming Event Management
Customer Focus	Purchasing and Inventory	Documents and Records	Assessments
Facilities and Safety	Equipment	Information Management	Continual Improvement

C24 covers the QSEs indicated by an "X." For a description of the other documents listed in the grid, please refer to the Related CLSI Reference Materials section.

Organization	Customer Focus	Facilities and Safety	Personnel	Purchasing and Inventory	Equipment	Process Management	Documents and Records	Information Management	Nonconforming Event Management	Assessments	Continual Improvement
		M29			X	X EP05 EP06 EP09 EP15 EP23 EP26 EP31					

Path of Workflow

A path of workflow is the description of the necessary processes to deliver the particular product or service that the organization or entity provides. A laboratory path of workflow consists of the sequential processes: preexamination, examination, and postexamination and their respective sequential subprocesses. All laboratories follow these processes to deliver the laboratory's services, namely quality laboratory information.

C24 covers the medical laboratory path of workflow step indicated by an "X." For a description of the other documents listed in the grid, please refer to the Related CLSI Reference Materials section.

Preexamination				**Examination**			**Postexamination**	
Examination ordering	Sample collection	Sample transport	Sample receipt and processing	Examination	Results review and follow-up	Interpretation	Results reporting and archiving	Sample management
EP31		EP31		X EP23	EP23 EP31	EP23		

Related CLSI Reference Materials*

EP05 **Evaluation of Precision of Quantitative Measurement Procedures. 3rd ed., 2014.** This document provides guidance for evaluating the precision performance of quantitative measurement procedures. It is intended for manufacturers of quantitative measurement procedures and for laboratories that develop or modify such procedures.

EP06 **Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach. 1st ed., 2003.** This document provides guidance for characterizing the linearity of a method during a method evaluation; for checking linearity as part of routine quality assurance; and for determining and stating a manufacturer's claim for linear range.

EP09 **Measurement Procedure Comparison and Bias Estimation Using Patient Samples. 3rd ed., 2013.** This document addresses the design of measurement procedure comparison experiments using patient samples and subsequent data analysis techniques used to determine the bias between two *in vitro* diagnostic measurement procedures.

EP15 **User Verification of Precision and Estimation of Bias. 3rd ed., 2014.** This document describes the estimation of imprecision and of bias for clinical laboratory quantitative measurement procedures using a protocol that can be completed within as few as five days.

EP23™ **Laboratory Quality Control Based on Risk Management. 1st ed., 2011.** This document provides guidance based on risk management for laboratories to develop quality control plans tailored to the particular combination of measuring system, laboratory setting, and clinical application of the test.

EP26 **User Evaluation of Between-Reagent Lot Variation. 1st ed., 2013.** This document provides guidance for laboratories on the evaluation of a new reagent lot, including a protocol using patient samples to detect significant changes from the current lot.

EP31 **Verification of Comparability of Patient Results Within One Health Care System. 1st ed., 2012.** This document provides guidance on how to verify comparability of quantitative laboratory results for individual patients within a health care system.

M29 **Protection of Laboratory Workers From Occupationally Acquired Infections. 4th ed., 2014.** Based on US regulations, this document provides guidance on the risk of transmission of infectious agents by aerosols, droplets, blood, and body substances in a laboratory setting; specific precautions for preventing the laboratory transmission of microbial infection from laboratory instruments and materials; and recommendations for the management of exposure to infectious agents.

* CLSI documents are continually reviewed and revised through the CLSI consensus process; therefore, readers should refer to the most current editions.

NOTES

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

Explore the Latest Offerings From CLSI!

As we continue to set the global standard for quality in laboratory testing, we are adding products and programs to bring even more value to our members and customers.

By becoming a CLSI member, your laboratory will join 1,600+ other influential organizations all working together to further CLSI's efforts to improve health care outcomes. You can play an active role in raising global laboratory testing standards—in your laboratory, and around the world.

Find out which membership option is best for you at **www.clsi.org/membership.**

Find what your laboratory needs to succeed! CLSI U provides convenient, cost-effective continuing education and training resources to help you advance your professional development. We have a variety of easy-to-use, online educational resources that make *e*Learning stress-free and convenient for you and your staff.

See our current educational offerings at **www.clsi.org/education.**

When laboratory testing quality is critical, standards are needed and there is no time to waste. eCLIPSE™ Ultimate Access, our cloud-based online portal of the complete library of CLSI standards, makes it easy to quickly find the CLSI resources you need.

Learn more and purchase eCLIPSE at **clsi.org/eCLIPSE.**

For more information, visit www.clsi.org today.

CLINICAL AND
LABORATORY
STANDARDS
INSTITUTE®

950 West Valley Road, Suite 2500, Wayne, PA 19087 USA

P: +1.610.688.0100 **Toll Free (US):** 877.447.1888 **F:** +1.610.688.0700

E: customerservice@clsi.org **www.clsi.org**

PRINT ISBN 1-56238-946-7

ELECTRONIC ISBN 1-56238-947-5