C54-A-IR
Vol. 32 No. 11
Replaces C54-A
Vol. 28 No. 19

Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision)

NOTE: Multiple corrections have been made to the formulae and information in this document. For a listing of all corrections, see page xi.

This document provides guidance on how to verify comparability of quantitative laboratory results for individual patients within a health care system.

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

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

ifcc International Federation of Clinical Chemistry and Laboratory Medicine

Clinical and Laboratory Standards Institute

Advancing Quality in Health Care Testing

Clinical and Laboratory Standards Institute (CLSI) is an international, interdisciplinary, nonprofit, standards developing, and educational organization that promotes the development and use of voluntary consensus standards and guidelines within the health care community. We are recognized worldwide for the application of our unique consensus process in the development of standards and guidelines for patient testing and related health care issues. Our process is based on the principle that consensus is an effective way to improve patient testing and health care services.

In addition to developing and promoting the use of voluntary consensus standards and guidelines, we provide an open and unbiased forum to address critical issues affecting the quality of patient testing and health care.

PUBLICATIONS

A document is published as a standard, guideline, or report.

Standard A document developed through the consensus process that clearly identifies specific, essential requirements for materials, methods, or practices for use in an unmodified form. A standard may, in addition, contain discretionary elements, which are clearly identified.

Guideline A document developed through the consensus process describing criteria for a general operating practice, procedure, or material for voluntary use. A guideline may be used as written or modified by the user to fit specific needs.

Report A document that has not been subjected to consensus review and is released by the appropriate consensus committee.

CONSENSUS PROCESS

CLSI's voluntary consensus process establishes formal criteria for the following:

- Authorization of a project
- Development and open review of documents
- Revision of documents in response to users' comments
- Acceptance of a document as a consensus standard or guideline

Invitation for Participation in the Consensus Process

Core to the development of all CLSI documents is the consensus process. Within the context and operation of CLSI, voluntary consensus is substantial agreement by materially affected, competent, and interested parties that may be obtained by following the consensus procedures defined in CLSI's Administrative Procedures. 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 are willing to accept the resulting agreement. CLSI documents are expected to undergo evaluation and modification in order to keep pace with advancements in technologies, procedures, methods, and protocols affecting the laboratory or health care.

Comments on Draft Documents

CLSI's voluntary consensus process depends on experts who 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. All comments along with the committee's responses are retained on file at CLSI and are available upon request.

Comments on Published Documents

The comments of users of published CLSI documents are essential to the consensus process. Anyone may submit a comment. All comments are addressed according to the consensus process by a committee of experts. A summary of comments and committee responses is retained on file at CLSI and is available upon request. Readers are strongly encouraged to comment at any time on any document.

APPEALS PROCESS

CLSI consensus procedures include an appeals process that is described in detail in the CLSI Administrative Procedures.

VOLUNTEER PARTICIPATION

Health care professionals in all specialties are urged to volunteer for participation in CLSI projects.

For further 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
610.688.0100
F: 610.688.0700
www.clsi.org
standard@clsi.org

Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision)

Abstract

Clinical and Laboratory Standards Institute document C54-A-IR—*Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision)* provides guidance on how to verify comparability of quantitative laboratory results for individual patients across a health care system. For the purpose of this document, a health care system is defined as a system of physician offices, clinics, hospitals, and reference laboratories, under one administrative entity, where a patient may present for laboratory testing, and whose results may be reviewed by any health care provider within the system for the purpose of providing medical care. This document does not provide guidance on how to correct method noncomparability that may be identified.

Clinical and Laboratory Standards Institute (CLSI). *Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision)*. CLSI document C54-A-IR (ISBN 1-56238-851-7 [Print]; ISBN 1-56238-852-5 [Electronic]). Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087 USA, 2012.

C54-A-IR
ISBN 1-56238-851-7 (Print)
ISBN 1-56238-852-5 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic)

Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision)

Volume 32 Number 11

Christopher M. Lehman, MD
John Rex Astles, PhD, FACB
Renze Bais, PhD
Sterling Bennett, MD
Ellis Jacobs, PhD, DABCC, FACB
Stan R. Johnson, MA
W. Gregory Miller, PhD
Jeffrey E. Vaks, PhD
Harvey B. Lipman, PhD
Amit Phansalkar, MS
Kenneth A. Sikaris, MD
Dietmar Stöckl, PhD
Greg Cooper, CLS, MHA

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

IFCC International Federation of Clinical Chemistry and Laboratory Medicine

Suggested Citation

CLSI. *Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision)*. CLSI document C54-A-IR. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.

Proposed Guideline
October 2007

Approved Guideline
May 2008

Approved Guideline (Interim Revision)
August 2012

ISBN 1-56238-851-7 (Print)
ISBN 1-56238-852-5 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic)

Committee Membership

The changes in this interim revision were approved by the Consensus Committee on Clinical Chemistry and Toxicology as follows:

David G. Grenache, PhD, DABCC, FACB
Chairholder
University of Utah, ARUP Laboratories
Salt Lake City, Utah, USA

Loralie J. Langman, PhD
Vice-Chairholder
Mayo Clinic
Rochester, Minnesota, USA

Julianne Cook Botelho, PhD
Centers for Disease Control and Prevention
Atlanta, Georgia, USA

Yung W. Chan, MT(ASCP)
FDA Center for Devices and Radiological Health
Rockville, Maryland, USA

Corinne R. Fantz, PhD, DABCC
Emory University
Atlanta, Georgia, USA

T. Scott Isbell, PhD, DABCC, FACB
Nova Biomedical Corporation
Chicago, Illinois, USA

Jessie Shih, PhD
Abbott
Abbott Park, Illinois, USA

Graham H. White, PhD
Flinders Medical Centre
Bedford Park, South Australia

Jack Zakowski, PhD, FACB
Beckman Coulter
Brea, California, USA

Staff

Clinical and Laboratory Standards Institute
Wayne, Pennsylvania, USA

Luann Ochs, MS
Senior Vice President – Operations

Ron S. Quicho
Staff Liaison

Megan P. Larrisey, MA
Editor

Ryan J. Torres
Assistant Editor

Acknowledgments

CLSI and the Consensus Committee on Clinical Chemistry and Toxicology gratefully acknowledge James Huntington and Simon Huntington, Co-founders, Analyse-it®, Leeds, United Kingdom, for their unwavering commitment and focused effort on the joint venture partnership with CLSI in the development of software to help laboratories easily implement these CLSI statistical methods.

Special thanks go to the following experts for carefully reviewing the statistics in C54-A-IR and applying their expert knowledge of statistical analysis for method validation to identify and offer solutions for the discrepancies and errors that have been corrected in this interim revision:

Jeffrey R. Budd, PhD
Beckman Coulter
Chaska, Minnesota, USA

Karl De Vore
Bio-Rad Laboratories, Inc.
Irvine, California, USA

Douglas M. Hawkins, PhD
University of Minnesota
Minneapolis, Minnesota, USA

Christopher M. Lehman, MD
University of Utah Health Sciences Center
Salt Lake City, Utah, USA

Curtis A. Parvin, PhD
Bio-Rad Laboratories
Plano, Texas, USA

The previous version of the document, C54-A (published in May 2008), was approved by the following CLSI committees:

Area Committee on Clinical Chemistry and Toxicology

David A. Armbruster, PhD, DABCC, FACB
Chairholder
Abbott Diagnostics
Abbott Park, Illinois

Christopher M. Lehman, MD
Vice-Chairholder
Univ. of Utah Health Sciences Center
Salt Lake City, Utah

John Rex Astles, PhD, FACB
Centers for Disease Control and Prevention
Atlanta, Georgia

David M. Bunk, PhD
National Institute of Standards and Technology
Gaithersburg, Maryland

David G. Grenache, PhD, MT(ASCP), DABCC
University of Utah
Salt Lake City, Utah

Steven C. Kazmierczak, PhD, DABCC, FACB
Oregon Health and Science University
Portland, Oregon

Linda Thienpont, PhD
University of Ghent
Ghent, Belgium

Jeffrey E. Vaks, PhD
Roche Molecular Diagnostics
Pleasanton, California

Hubert Vesper, PhD
Centers for Disease Control and Prevention
Atlanta, Georgia

Jack Zakowski, PhD, FACB
Beckman Coulter, Inc.
Brea, California

Advisors

Mary F. Burritt, PhD
Mayo Clinic
Scottsdale, Arizona

Paul D'Orazio, PhD
Instrumentation Laboratory
Lexington, Massachusetts

Carl C. Garber, PhD, FACB
Quest Diagnostics, Incorporated
Lyndhurst, New Jersey

Uttam Garg, PhD, DABCC
Children's Mercy Hospitals & Clinics
Kansas City, Missouri

Neil Greenberg, PhD
Ortho-Clinical Diagnostics, Inc.
Rochester, New York

Harvey W. Kaufman, MD
Quest Diagnostics, Incorporated
Lyndhurst, New Jersey

W. Gregory Miller, PhD
Virginia Commonwealth University
Richmond, Virginia

Gary L. Myers, PhD
Centers for Disease Control and Prevention
Atlanta, Georgia

David Sacks, MD
Brigham and Women's Hospital and Harvard Medical School
Boston, Massachusetts

Bette Seamonds, PhD
Mercy Health Laboratory
Swarthmore, Pennsylvania

Dietmar Stöckl, PhD
STT Consulting
Horebeke, Belgium

Thomas L. Williams, MD
Nebraska Methodist Hospital
Omaha, Nebraska

Subcommittee on Verification of Comparability of Patient Results

Christopher M. Lehman, MD
Chairholder
Univ. of Utah Health Sciences Center
Salt Lake City, Utah

John Rex Astles, PhD, FACB
Centers for Disease Control and Prevention
Atlanta, Georgia

Renze Bais, PhD
Pacific Laboratory Medicine Services
Sydney, Australia

Sterling Bennett, MD
LDS Hospital
Salt Lake City, Utah

Ellis Jacobs, PhD, DABCC, FACB
NYU/Bellevue
New York, New York

Stan R. Johnson, MA
Beckman Coulter, Inc.
Brea, California

W. Gregory Miller, PhD
Virginia Commonwealth University
Richmond, Virginia

Jeffrey E. Vaks, PhD
Roche Molecular Diagnostics
Pleasanton, California

Ian S. Young, MD, FRCP
Queen's University Belfast
Belfast, United Kingdom

Advisors

J. David Bessman, MD
Univ. of Texas Medical Branch
Galveston, Texas

Elma Kamari Bidkorpeh
Kaiser Permanente
North Hollywood, California

Harvey B. Lipman, PhD
Centers for Disease Control and Prevention
Atlanta, Georgia

Amit Phansalkar, MS
ARUP Laboratories
Salt Lake City, Utah

Acknowledgments

This guideline was prepared by CLSI, as part of a cooperative effort with IFCC to work toward the advancement and dissemination of laboratory standards on a worldwide basis. CLSI gratefully acknowledges the participation of IFCC experts Ian S. Young, MD, FRCP; and Renze Bais, PhD, on this project.

In addition, CLSI and the Subcommittee on Verification of Comparability of Patient Results gratefully acknowledge the following volunteer for his important contributions to the development and/or completion of this document: Greg Cooper, CLS, MHA, Bio-Rad Laboratories, Inc.

Contents

Contents (Continued)

Interim Revision Changes to C54-A

Section 4

- A definition has been added for "standard deviation."

Sections 8.2.1

- An explanation of the "range test" has been substituted for the "studentized range test" description.

Section 10.3

- A modified protocol for identifying an approximate analyte concentration to be used for the range test has been substituted. The new protocol requires knowledge of both total and within-run precision of the measurement system at the selected analyte concentration.

Section 10.7

- A description of how to calculate the critical difference for the range test has been included.

Sections 10.8

- A protocol for determining both the number of runs and number of replicates per run for the range test has been included.

Appendix A. Worked Examples

- Examples of how to use the document have been revised.

Appendix B. Tables of Runs, Replicates, and Range Rejection Limits

- New tables for use in determining both the number of runs and number of replicates per run for the range test have been included, as well as a description of how to use the tables.

Appendix C. Statistical Concepts

- C4. Range Test—The mathematical basis for the range test replaces Section C4 of the prior version of the document.

- C5. Within-run vs Total Standard Deviation—A description of the components of the precision of a measurement system are provided to elucidate the need for two runs when between-run imprecision makes up a significant proportion of total SD. This replaces Section C5 of the prior version of the document.

- C6. Number of Replicates—This section was eliminated.

- C7. Power Curves—This section was eliminated.

- C8. Comparative Power of Test Procedures—This section was eliminated.

Summary of Consensus and Delegate Comments and Subcommittee Responses

- The Summary of Consensus and Delegate Comments and Subcommittee Responses was removed as part of this interim revision. This summary is on file at the CLSI office and available upon request by contacting CLSI at 610.688.0100 or standard@clsi.org.

Foreword

Patients may present for laboratory testing at multiple locations within a health care system. Continuity of medical care requires that the comparability of test results produced by different measurement systems be verified periodically. This document provides guidance on how to verify the comparability of quantitative laboratory results for analytes tested on different measurement systems. The document addresses causes of noncomparability, risk assessment of comparability failure, frequency of comparison testing, concentrations to be compared, commutability of comparability testing materials, a comparability testing protocol, and acceptance criteria for interpretation of comparability testing. The comparability testing protocol described in this document is an intuitive, simple approach that balances the need for a statistically valid, clinically relevant methodology with practical limitations on laboratory resources. Other valid procedures for comparability evaluation can be developed by a laboratory, and it is not the intent of this document to exclude their use. This protocol can also be used to validate reagent lot changes.

Key Words

Accuracy, bias, coefficient of variation, commutability, comparability, imprecision, range test

Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision)

1 Scope

This document provides guidance on how to verify comparability of quantitative laboratory results for individual patients within a health care system. For the purpose of this document, a health care system is defined as a system of physician offices, clinics, hospitals, and reference laboratories, under one administrative entity, where a patient may present for laboratory testing, and whose results may be reviewed by any health care provider within the system for the purpose of providing medical care.

C54 provides a simple approach to be used for the assessment of patient laboratory result comparability across a maximum of 10 instruments, and assumes that a more comprehensive validation of quantitative measurement system comparability has been undertaken when the measurement systems were initially introduced into the laboratory. A more comprehensive comparison among measurement procedure results can follow a methodology such as that described in CLSI document EP09.[1] Comparability testing is just one facet of a program for assuring quality laboratory performance and is not intended to be a substitute for other quality monitors. This document does not address corrective action should method noncomparability be identified.

The approach described can also be used to verify comparability of patients' results in situations such as those following reagent or calibrator lot changes, instrument component changes or maintenance procedures, alerts from QC or external quality assessment (EQA) (proficiency testing [PT]) events, or other special cause event.

2 Introduction

Out of necessity, or for their own convenience, patients may interface with health care systems for the purpose of laboratory testing in a variety of settings and/or locations. Results of these tests may be compiled and reviewed by providing clinicians at any of the patient care locations. In addition, larger laboratories may have multiple instruments within one location (eg, backup instruments, point-of-care [POC] instruments) that may provide laboratory results for an individual patient during a health care episode. Over time, lots of calibrator and reagents change, calibration and maintenance procedures are performed, and other events may occur that can affect patient test results. The diagnostic value of patient test results is maximized if the measurement systems providing such results are in a state of statistical control (ie, are producing stable and consistent results). Maintaining comparability may involve standardization and calibration of instruments, forced agreement of results among different measurement systems through mathematical transformation, or adoption of different reference intervals and/or therapeutic or diagnostic cutoffs that are clearly indicated in the patient report. Regardless of the approach used to achieve comparable results among different measurement systems, or to accommodate known differences, periodic verification of assay comparability is necessary to provide optimal patient care.

There is no consensus procedure for demonstrating patient laboratory result comparability for patient samples among measurement procedures. A survey of the participants involved in the preparation of this document demonstrated a variety of approaches to testing frequency, number and type of samples tested (eg, random, high and low concentrations, or concentrations spanning the analytical measurement range [AMR]), evaluation and acceptance criteria for the results of comparison testing, and method of dealing with known bias between methods. The intent of this document is to review the salient issues surrounding verification of comparability of patient results among measurement procedures, and to provide a practical, statistically valid approach that laboratories of varying size and resources can use to satisfy this quality

requirement. Other valid procedures for comparability evaluation can be developed by a laboratory, and it is not the intent of this document to exclude their use.

This guideline addresses evaluation and monitoring of comparability of patient results. Recommendations on monitoring stability of the analytical process are provided in CLSI document C24.[2] Other clinical laboratory procedures are in place to address calibration traceability of routine measurement procedures to reference systems that are intended to ensure long-term consistency of calibration and uniformity of results among providers of *in vitro* diagnostic (IVD) measurement systems (see CLSI document X05[3] and ISO 17511[4] for further information).

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 blood-borne pathogens. The Centers for Disease Control and Prevention address this topic in published guidelines that focus on the daily operations of diagnostic medicine in human and animal medicine while encouraging a culture of safety in the laboratory.[5] 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 disease, refer to CLSI document M29.[6]

4 Terminology

4.1 Definitions

accuracy (measurement) – closeness of agreement between a measured quantity value and a true quantity value of a measurand (JCGM 200:2012).[7]

alpha error – probability of falsely rejecting the null hypothesis when it is true.

analyte – component represented in the name of a measurable quantity (ISO 17511).[4]

analytical measurement range (AMR) – the range of analyte values that a method can directly measure on the sample without any dilution, concentration, or other pretreatment that is not part of the typical assay process.

beta error – probability of falsely rejecting the alternative hypothesis when it is true.

bias – difference between the expectation of the test results and an accepted reference value (ISO 5725-1,[8] ISO 3534-1[9]); **NOTE 1:** Bias is the total systematic error, as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value (ISO 5725-1)[8]; **NOTE 2:** The measure of trueness is usually expressed in terms of bias (ISO 3534-1).[9]

calibration – operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication (JCGM 200:2012).[7]

calibrator – substance, material, or article intended by its manufacturer to be used to establish the measurement relationships of an *in vitro* diagnostic medical device.

coefficient of variation (CV) – for a non-negative characteristic, the ratio of the SD to the average (ISO 3534-1)[9]; **NOTE:** The ratio may be expressed as a percentage.

commutable – interassay properties of a reference material, calibrator material, or QC material that are comparable with those demonstrated by authentic clinical specimens; **NOTE:** Commutability of a material is defined as the "degree to which a material yields the same numerical relationships between results of measurements by a given set of measurement procedures, purporting to measure the same quantity, as those between the expectations of the relationships obtained when the same procedures are applied to other relevant types of material" (ISO 15194).[10]

comparability – agreement between patient results obtained for an analyte using different measurement procedures within a health care system; **NOTE:** The results are considered to be comparable if the differences do not exceed a critical value established based on defined acceptance criteria.

external quality assessment (EQA)//proficiency testing (PT) – 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.

imprecision – the random dispersion of a set of replicate measurements and/or values expressed quantitatively by a statistic, such as SD or CV; **NOTE:** It is defined in terms of repeatability and reproducibility.

measurand – quantity intended to be measured (JCGM 200:2012).[7]

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 (JCGM 200:2012)[7]; **NOTE 1:** A measurement procedure is usually documented in sufficient detail to enable an operator to perform a measurement (JCGM 200:2012)[7]; **NOTE 2:** This term pertains to specific procedures as marketed by specific manufacturers; **NOTE 3:** In other documents, equivalent terms were *method, device,* and *assay;* **NOTE 4:** A measurement procedure is based on a measurement method.

measurement system – a unit or device used to measure or assess the presence or absence of a particular substance, or to quantitate that substance, found in blood or body fluids; **NOTE:** A measurement system includes instructions and all of the instrumentation, equipment, reagents, and/or supplies needed to perform an assay or examination and generate test results.

point-of-care testing (POCT)//bedside, near-patient testing – testing performed in an alternate site, outside a central laboratory environment, generally nearer to, or at the site of, the patient.

power – probability of accepting the alternative hypothesis that a substance causes interference when it is true; **NOTE:** The probability is usually denoted as a percentage, $100(1-\beta)\%$.

precision (of measurement) – closeness of agreement between independent test results obtained under stipulated conditions (ISO 3534-1)[9]; **NOTE:** The measure of precision usually is expressed in terms of imprecision and computed as an SD of the test results. Less precision is reflected by a larger SD (ISO 3534-1).[9]

proficiency testing (PT)//external quality assessment (EQA) – determination of laboratory testing performance by means of interlaboratory comparisons; **NOTE 1:** Commonly, a program periodically sends multiple specimens to members of a group of laboratories for analysis and/or identification; the program then compares each laboratory's results with those of other laboratories in the group and/or with an assigned value, and reports the results to the participating laboratory and others; **NOTE 2:** Other forms of PT/EQA include: data transformation exercises, single-item testing (in which one item is sent to a number of laboratories sequentially and returned to the program at intervals), and one-off exercises (in which laboratories are provided with a test item on a single occasion); **NOTE 3:** The results are summarized, analyzed, and, with some tests, graded by the program and provided to the participating site, which can compare its results with those of other sites that use a similar method.

quality control (QC) – the operational techniques and activities that are used to fulfill requirements for quality (ISO 9000)[11]; **NOTE 1:** *In health care testing,* the set of procedures designed to monitor the test method and the results to assure test system performance; **NOTE 2:** QC includes testing control 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.

risk – combination of the probability of occurrence of harm and the severity of that harm (ISO 15190[12]; ISO/IEC Guide 51[13]).

sample – one or more parts taken from a system, and intended to provide information on the system, often to serve as a basis for decision on the system or its production (ISO 15189)[14]; **NOTE 1:** For example, a volume of serum taken from a larger volume of serum (ISO 15189)[14]; **NOTE 2:** A sample is prepared from the patient specimen and used to obtain information by means of a specific laboratory test; **NOTE 3:** For the purposes of this guideline, readers can consider the terms "sample" and "specimen" to be equivalent; **NOTE 4:** The term "specimen" has been used in laboratory medicine as a synonym for a sample, as defined here, of biological origin, or for an entire macroscopic parasite.

standard deviation (SD) – a measure of variability/dispersion that is the positive square root of the population variance.

statistic – a function of a set of observations from a random variable; **NOTE:** A statistic is also a random variable; thus, it also has statistics, such as mean and SD.

total error – the sum of any set of defined errors that can affect the accuracy of an analytical result; **NOTE:** Total error can be defined as the sum of bias and imprecision.

trueness (of measurement) – closeness of agreement between the average value obtained from a large series of test results and an accepted reference value (ISO 3534-1)[9]; **NOTE:** The measure of trueness is usually expressed in terms of bias (ISO 3534-1).[9]

Type I error – an incorrect judgment or conclusion that occurs when an association is found between variables where, in fact, no association exists; **NOTE 1:** For example, if the experimental procedure does not really have any effect, chance or random error may cause the researcher to conclude that the experimental procedure did have an effect; **NOTE 2:** Also known as "false positive" or "alpha error."

Type II error – an incorrect judgment or conclusion that occurs when no association is found between variables where, in fact, an association does exist; **NOTE 1:** In a medical screening, for example, a negative test result may occur by chance in a subject who possesses the attribute for which the test is conducted; **NOTE 2:** Also known as "false negative" or "beta error."

validation – confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled (ISO 9000).[11]

4.2 Abbreviations and Acronyms

AMR	analytical measurement range
AST	aspartate transaminase
CD	critical difference
CDF	cumulative distribution function
CV	coefficient of variation
CV_G	between-subject biological variability
CV_I	within-subject biological variability
EQA	external quality assessment
HbA1c	hemoglobin A1c
IVD	*in vitro* diagnostic
ISO	International Organization for Standardization
POC	point-of-care
POCT	point-of-care testing
PT	proficiency testing
PT	prothrombin time
QA	quality assurance
QC	quality control
SD	standard deviation
SR	within-run standard deviation
ST	total standard deviation
WBC	white blood cell

5 Practical Considerations for Designing a Comparability Monitoring Protocol

A number of factors should be considered when designing a comparability protocol. The laboratory director must determine the appropriate protocol for monitoring each analyte that is measured by more than one instrument in the health care system. Applicable regulatory and/or accreditation requirements (eg, frequency of comparison testing) should be incorporated into the design of any protocol. Comparability verification is considered good laboratory practice even if it is not a regulatory or accreditation requirement.

5.1 Causes of Noncomparability of Results

In designing a plan for routine assessment of measurement system comparability, potential causes of noncomparability of results for patients' samples should be considered. Reasons for differences between results from more than one instrument or method include:

- Different analytical methodologies
- Differences in calibration between measurement procedures
- Differences in imprecision between measurement procedures
- Existence of value assignment errors and variation of commutability between lots of calibrators
- Simultaneous use of calibrator lots of different ages/stages of time-dependent degradation in different laboratory locations
- Differences in commutability of calibrators with different measurement procedures from different IVD manufacturers

- Reagent on-instrument degradation after calibration
- Instrument drift/failure
- Use of different reagent lots, or differences in packaging, shipping, or storage conditions when the same method is used on more than one instrument
- Differences in instrument analytical parameters, such as dilution ratios and incubation times between different instruments that use the same reagents
- Preexamination effects on the sample, including differences in sample handling between different types of measurement systems

Differences due to calibration, reagent lots, and instrument parameters are more easily managed by the laboratory to achieve comparable results for patient samples, while differences between results caused by fundamental differences in analytical methodology are more difficult to manage. For example, antibodies may be directed against different epitopes of a polypeptide hormone, in which case the substance actually measured is different depending upon the method. When there are fundamental differences in analytical methodologies used within a health care system, it may be impossible to force patient results to agree through a calibration process or by adjusting reported results using a mathematical correction factor. In general, such differences are more frequently addressed by defining and reporting different reference intervals.

5.2 Scope of Comparisons

Ideally, every measurement system measuring the same analyte in the health care system should be included in comparability testing. For unstable analytes, it may be possible to stabilize patient specimens collected for comparability testing (eg, glucose may be stabilized in whole blood with fluoride ion, and ammonia may be stabilized by freezing plasma aliquots). In exceptional cases, materials other than patient specimens may be required for comparability testing (see Section 6).

5.3 Risk Assessment for Noncomparable Results

For comparability testing, risk equals the product of the potential for harm caused by the degree of measurement system noncomparability and the potential frequency of occurrence of noncomparable results for a specific analyte.

5.3.1 Clinical Impact of Noncomparability

The laboratory director must assess the impact of noncomparability of a measurement system on patient care. Input from practicing clinicians who order the test should be solicited when necessary. Harm to the patient may result when diagnosis and/or treatment is delayed due to clinician confusion about noncomparable results that may generate an additional “tie-breaker” test (eg, noncomparability of emergency department and outpatient clinic results). In addition, physicians directing the care of patients who are being monitored for various biomarker or therapeutic drug levels, which are measured on instruments that give dissimilar results, may be confused about the outcomes of treatments or dosing regimens (eg, tumor markers). Analytes that are at risk for noncomparability and that pose a significant risk to patient outcome may warrant more frequent and/or more rigorous assessment.

5.3.2 Probability of Noncomparability

The laboratory director must assess the likelihood that two assays will demonstrate noncomparability given the inherent limitations of the measurement systems being compared (see Section 7). Evidence of

calibration instability or prior problems maintaining comparability, for example, may indicate a need for more frequent comparisons, and/or more frequent calibrations.

5.4 Frequency and Complexity of Comparability Assessment Protocols

In designing a comparability protocol for an analyte, the laboratory director must consider the risk to patients of noncomparability of assays, as well as practical considerations. Approaches to comparability testing can vary significantly in terms of reagents consumed; time spent procuring, storing, transporting, and measuring samples for comparative analysis; and time spent evaluating the comparability of results. It may be useful to begin monitoring comparability with as high a frequency of comparisons as indicated by risk assessment and cost effectiveness, making improvements based on comparison data. The frequency of monitoring can then be reduced based on improvements in performance and revisions in risk assessment (if applicable).

5.4.1 Statistical Considerations

From a statistical standpoint, the tolerance for falsely detecting a difference between assays (Type I error) must be balanced against the tolerance for failure to detect a true difference (Type II error) that is clinically significant. Practical considerations frequently limit the sample size of a comparison, increasing the probability of a Type II error. In addition, frequent comparisons of stable assays increase the probability of a Type I error (see Section 8 and Appendix C).

5.4.2 Operational and Cost Considerations

There are a number of practical considerations that the laboratory director should address when deciding how frequently to verify comparability and the number of replicates to be tested. Operational factors that may influence the frequency of testing include staffing availability (staff may need to spend considerable time acquiring, storing, and transporting specimens); availability and stability of appropriate samples for testing; capacity for storing patient specimens; geographic locations of testing sites; cost of reagents; and the opportunity to combine comparability testing with other QA testing, such as verification of the AMR (however, see discussion of appropriate sample selection in Section 6). Ultimately, the laboratory director must balance the risk to the patient of noncomparable results against the cost in materials and labor to the laboratory when designing a protocol for evaluation and maintenance of interassay comparability. If the monitoring of comparability is accompanied by process improvements, the initial cost of implementing comparability monitoring and the cost of subsequent process improvements should be mitigated by cost savings due to improved performance and cost reductions due to less frequent monitoring.

5.5 General Approaches to Comparability Testing

Interinstrument comparability testing can be categorized as the following:

- Frequent (eg, daily, weekly) monitoring
- Periodic monitoring (eg, quarterly, biannually) that is performed when frequent monitoring is deemed unnecessary because the measurement systems involved are stable and the risk of errors in clinical interpretation due to noncomparable results is low
- Special cause testing that is performed in response to an alert from a monitoring procedure or other triggering event (see Section 5.6) when a greater degree of statistical confidence in the results is desired.

Frequent monitoring may be set up to have more or less power to detect a difference, depending upon the requirements of the assay, but generally involves comparing fewer samples or running fewer replicates of

a single specimen. This approach is relatively low-cost in terms of number of patient samples tested, and time and reagents consumed per comparison event. Alternatively, frequent monitoring can be accomplished through automated, statistical monitoring of patient results (eg, weighted moving averages).[15] Frequent monitoring provides the opportunity to evaluate trends in comparability of results over time, and allows for better understanding and improvement of the measurement procedure. Periodic monitoring should be designed to have greater power to detect a difference (ie, a larger number of patient samples or replicate analysis of individual samples) due to the lower frequency of comparisons. Consequently, periodic monitoring is generally more costly in terms of time and reagents per comparison event. Special cause testing often requires a larger number of patient samples or replicate testing of individual samples to provide increased statistical power to detect a clinically significant difference between assays. Special cause testing should be used for troubleshooting and follow-up to resolve comparability issues identified by a monitoring procedure. It is important to note that although sample sizes will be larger than periodic monitoring, sample sizes for special cause assessments are generally expected to be smaller than what is required for an initial method validation.

Comparability testing should only be conducted when all measurement systems that are being compared are functioning according to the manufacturer's recommendations and are judged to be in control. However, comparability testing is a component of the QA process and may provide an indication that a measurement procedure needs to be reviewed for possible corrective action.

5.6 Triggers for Special Cause Comparability Testing

5.6.1 Failure of a Frequent or Periodic Monitor

When frequent or periodic comparison testing fails to pass acceptance criteria, it may be appropriate to follow up with special cause testing to confirm noncomparability and to document conformance after analytical issues have been resolved.

5.6.2 Proficiency Testing/External Quality Assessment Failure

Comparability testing among methods or instruments may be useful to investigate and resolve a PT/EQA failure. After correcting the analytical source of the PT/EQA failure, repeat testing between instruments may be necessary to confirm comparability.

5.6.3 Shift in a Statistical Monitoring Parameter

Hematology instruments generally have a built-in software feature that provides a weighted moving average of patient results for various parameters. Some chemistry instruments and laboratory information systems offer similar capability using various statistical procedures. If a change in the composition of the patient population being monitored has been ruled out, a shift in the moving average, or other statistical trend test, may be an indication for comparability testing, once any analytical issues have been resolved.

5.6.4 Quality Control Result Failure

Results produced from the analysis of QC samples are used to monitor and verify that a measurement system is performing within expectations for a stable measurement process. QC result acceptance criteria are designed to detect unacceptable imprecision and bias that exceed the expectations for stable measurement system performance. An unacceptable QC result or trend of results may be an indication for follow-up with patient sample–based comparability testing. It is important to note that QC materials are manufactured to simulate properties of patient samples, but the processing required to produce QC materials may cause them to be noncommutable with native clinical samples (see Section 6.1.3). Consequently, in most situations, the results for QC samples cannot be reliably compared between different instruments and methods as a surrogate for the comparability of patient results. However, when

identical analyzers—being monitored with identical QC materials—produce QC results that begin to deviate from each other, that may suggest the need for comparability testing.

5.6.5 Reagent or Calibrator Lot Change

Reagent and calibrator lot changes are a commonly occurring special cause for comparability testing. Good laboratory practice includes verification that patient results are comparable to those from a previous lot when a new lot of reagents or a new lot of calibrator is put into service. In some countries, regulatory or accreditation requirements dictate verification of performance following reagent or calibrator lot changes.

5.6.5.1 Reagent Lot Change

The principal consideration when introducing a new reagent lot is the choice of material to use to verify comparability of patient results with those from the prior lot. QC materials have been used for this purpose, but have inherent commutability limitations that may confound conclusions based on results following reagent lot changes. QC materials may have a different commutability characteristic (causing a different noncommutability bias) between two reagent lots, which can cause an apparent difference inconsistent with results for native patient samples, or can cause an apparent agreement or "false-negative result" when a real difference exists for native patient samples (see Section 6 for recommendations on materials to use for comparability testing, and limitations and verification practices necessary when using QC or other materials with unknown commutability properties).

5.6.5.2 Calibrator Lot Change

If a calibrator lot is changed at the same time as a reagent lot change, then the precautions in Section 5.6.5.1 for reagents are applicable.

If a calibrator lot change is made and the same reagent lot(s) continues in use, then the choice of materials to use for comparability testing is simplified because QC materials can be used without additional qualification. Commutability is a property of a non-native sample material that exists between a particular material and reagent combination. When there is no change in reagent lot, there is no change in the commutability property for a given material, and differences in results for a QC material between a new vs a prior lot of calibrator are expected to reflect the relationship for native patient samples.

5.6.6 Ad Hoc Comparability Testing

Ad hoc verification of comparability may be indicated in the following circumstances: following resolution of an underlying problem in one or more instruments, major maintenance, component replacement, software update, or clinician inquiry regarding the accuracy of results.

6 Samples for Comparability Testing

6.1 Commutability

The selection of materials for comparability testing should take into account the commutability of the material. Commutability is the equivalence of the mathematical relationships among the results of different measurement procedures for a reference material and for representative samples of the type intended to be measured. Freshly obtained patient samples represent the "ideal" material for comparison testing, because they are the intended samples to be analyzed by measurement systems under routine circumstances. When any other type of sample is used, its commutability with native patient samples must be verified. Commercial materials intended for use in routine QC, demonstration of linearity, external PT, and instrument calibration have, in most cases, had their matrices modified in ways that may

significantly affect commutability with native clinical samples.[16,17] A manufacturer's product calibrators are typically intended only for use with a specific routine measurement procedure and are not commutable for use with other manufacturers' measurement procedures.

6.1.1 Patient Specimens

The optimal samples for comparability testing are native patient samples collected in an appropriate, validated collection container and processed (and stored, if necessary) according to the stability requirements of the analyte. Samples containing substances known to interfere with the assay being compared should be excluded, because the purpose of the comparison testing in this guideline is to verify comparability of results for typical samples, not to verify the specificity of the methods. Because routine measurement procedures may not be completely specific for the analyte, there will be a distribution of results for native patient samples when measured by two or more procedures. The distribution of results will have contributions from the imprecision of measurement due to reproducibility and repeatability, and from sample-specific nonspecificity effects. The statistical criteria to determine if two or more routine measurement procedures have equivalent results should consider both sources of variability. However, as noted previously, if two different methods have different specificities for the analyte, it may not be possible to achieve comparable results for patient samples.

A second generally acceptable material for comparability testing is a pool of patient samples. Pooled samples should be used when the number of measurement systems to be compared requires more sample than is available from a single phlebotomy collection at one point in time. Pooled samples have the limitation of not adequately representing individual samples, because differences between individual patient samples may be masked. Interactions among donor samples may cause precipitation of some serum proteins and protein-bound molecules that are important as analytes or as potential interferents in the measurement procedures. Pooling may also dilute unspecified interfering substances to levels at which they no longer interfere with a method. Thus, pooling native samples may be an advantage for the purpose of comparability testing intended primarily to evaluate calibration differences (bias) among measurement systems.

Collecting and processing samples for preparation of a pool require careful consideration of analyte stability on storage before pooling and during the pooling process. CLSI document C37[18] includes information on handling blood and serum for preparing large pools of reference material, but the principles are applicable to smaller pools that would be used for comparability evaluation in one health care system.

It is often difficult for laboratories to locate patient samples containing analyte concentrations of interest for comparability testing. It may be necessary to add purified or partially purified analyte to native clinical samples or pools of native samples to achieve higher levels of an analyte. The additive, or an impurity in the additive, may have an unexpected influence on the matrix that would compromise the "native" characteristics of the resultant sample. For some analytes that are normally not detectable in healthy individuals, it may be possible to add a small amount of a sample with a very high amount of an analyte (eg, human chorionic gonadotropin) to a pool. In some cases, there may be a metabolite of an analyte that is also measured in a measurement procedure, in which case adding only unmetabolized analyte to a pool may not be appropriate for comparison testing. For example, a comparison of a gas chromatography mass spectrometry method with an immunoassay for measurement of cyclosporin would appear to demonstrate better comparability between methods for a sample that was spiked with pure drug than for native patient samples that contain a combination of drug and drug metabolites.

6.1.2 Commutable Reference and Control Materials

Reference materials, control materials, and PT/EQA materials that have been demonstrated to be commutable with patient specimens for the method(s) being compared are suitable for comparability testing.

6.1.3 Quality Control Materials

Under some circumstances, QC results may be used to verify comparability of results among different instruments and methods. When identical instruments use the same lot of reagents, there is a good probability that the relationship between the QC material results on each instrument will be very similar to the relationship for patient samples. This situation occurs because any matrix biases associated with noncommutable QC sample results are expected to be the same when identical instruments and reagents are used. However, when there are differences in instrument platforms (even from the same manufacturer), different lots of reagents are used (even on the same instrument model), or different measurement systems are used, it becomes increasingly likely that results produced from the analysis of QC materials will not have the same numerical relationship among methods as do results from native clinical samples, and erroneous conclusions regarding the comparability of patient sample results may be made.

Because QC materials are frequently noncommutable with native clinical samples, QC results may give an apparent numeric relationship between measurement systems that does not reflect the true relationship observed for patient samples. The lack of a difference observed using QC samples can be a "false negative," in which case a true difference for patient samples may be masked by an offsetting matrix bias that gives the false impression that the measurement systems produce comparable results. Conversely, an apparent difference observed using QC samples may be due only to a matrix bias, and the results for patient samples may in fact be equivalent between the instruments.

6.1.3.1 Using Quality Control Results From Different Measurement Systems

For a situation in which different instruments and/or reagents are used to measure the same analyte, the relationship between results for QC samples from different measurement systems can be trended to determine if any changes in the relationship have occurred. When a measuring system is performing within the expectations for a stable measurement process, the QC results should be stable and consistent for a given instrument/reagent combination. If the numeric relationship between the means for the QC results from two, or more, measuring systems is known, and the results for native clinical samples have been verified to be comparable between those measuring systems, the numeric relationship between the means of the QC results should remain constant as long as the performance of the measuring systems remains unchanged. The numeric relationship between the means for the QC results can be monitored over time (moving means) as an indicator that the previously established comparability based on native clinical samples has remained unchanged. However, the numeric relationship between the moving means for QC results may change every time the lot of QC material changes and every time the lot of reagents changes on any of the measuring systems. Consequently, it is necessary to reestablish the numeric relationship between means of the QC results, with verification that results for native clinical samples have remained comparable among the measurement systems following lot changes. It may be more practical to perform frequent or periodic comparability monitoring using native patient samples than to conduct the validation necessary to base monitoring on QC results.

6.1.3.2 Use of Quality Control Materials When Preexamination Stability of Patient Samples Is Limited

There may be circumstances when the preexamination stability of an analyte is a limiting factor for comparison testing (ie, stability is less than the time required to transport a sample aliquot to each of the instruments to be compared). There may also be situations when analyte concentration or activity at appropriate levels for testing cannot be realistically achieved with patient samples. In these cases, QC materials (or other reference or PT/EQA materials) may be the only samples available for comparison testing. The commutability limitations of QC, PT/EQA, and reference materials described above must be considered when making conclusions regarding the comparability of results for patient samples. Differences between measuring systems observed using noncommutable samples may be due to a true bias, a bias caused by the sample matrix, or a combination of both. The QC, reference, or PT/EQA

materials used should be validated for suitability in evaluating comparability of patient sample results for the measurement systems involved.

6.1.3.3 Use of Quality Control Materials, or Verification Materials, Provided by the Measurement System Manufacturer

An IVD manufacturer may provide control materials specifically designed and validated to verify that the performance of their measurement systems meets the claims in the product labeling. Although these materials may not be commutable with patient samples, the manufacturer may have designed the materials to have approximately constant performance among the measurement systems identified in the product labeling. Such materials may be used as samples for comparability testing among the measurement systems that are specifically identified in the product labeling. Review of the claims for such control materials and confirmation that the comparability limits meet the laboratory's clinical requirements are recommended.

A control material provided by one measurement system manufacturer will not be suitable for use with a measurement system from another manufacturer, because the material will not have been validated to be commutable with patient samples among the different measurement systems.

6.1.4 Materials Used for External Quality Assessment, Proficiency Testing, or External Group Quality Control Evaluation

If preexamination sample stability is a limiting factor such that patient samples cannot be used, materials used for external evaluation of performance may be considered for use as comparison materials. These types of materials are typically not validated for commutability with native clinical samples. However, the materials are typically analyzed by a large number of laboratories using the same instruments and methods, and the mean value within a group of peer instruments and methods will include bias components attributable to calibration and to noncommutability. However, the noncommutability can be assumed to be approximately constant within the peer group. The mean value for a peer group with a sufficiently large number of participants (usually considered to be ≥ 10), and an acceptable among-participant SD,[19,20] can be assumed to represent a value consistent with use of the measurement system in compliance with the manufacturer's instructions. However, there may be reagent lot–specific matrix biases within the peer group data that require wider acceptance limits than would be applicable for native patient samples.

An individual laboratory can use the mean value for the appropriate peer group to determine that an instrument/method combination has remained stable and continues to meet the performance verified at the time of the external assessment event. Many assessment programs allow participants to purchase extra vials of the materials to use for internal verification procedures. It is also possible to store residual quantities of the external assessment materials under storage conditions (usually frozen at −70°C) that will prevent degradation of the analyte. However, caution must be used when storing extra vials, or when storing and reusing residual external assessment materials, because they may not have been validated for this purpose. In particular for residual material, there may be deterioration of the analyte during its open vial use period or during storage, or deterioration caused by a freeze-thaw cycle; in addition, the matrix may be altered by the storage conditions.

The results from the analysis of an external assessment material cannot be used to directly compare different measurement procedures (ie, those that are not considered members of the same peer group). The material is not likely to be commutable with native clinical samples, and it is not possible to determine if the numeric relationship between different measurement procedures is influenced by the presence of a matrix bias (see Section 6.1.3). Consequently, results that appear to agree could have a calibration bias that is offset by a matrix bias.

6.1.5 Other Nonpatient Materials

Linearity verification materials and routine measurement calibration materials (ie, manufacturer's product calibrators) are not recommended for verification of comparability of patients' results, because these materials are not intended to be commutable with native patient samples.

6.2 Analyte Concentrations for Testing

The laboratory director must consider the clinical use of an assay and practical limitations on the laboratory when designing comparison testing procedures. To apply the comparability method described in this document, comparisons must be made at analyte concentrations where there is a reliable estimate of the imprecision of measurement (see Section 8.2.2). Most laboratories have insufficient resources to test across the full AMR of a measurement system at each comparison testing event, although this would be the ideal approach. Laboratories may choose to perform comparability testing near the mean concentration(s) of QC material(s), because imprecision is known at these concentrations; near significant clinical decision values or upper and lower reference interval limits; or some combination of these across consecutive comparability testing events.

6.3 Storage and Transport

A laboratory may choose to store native patient samples containing specific analyte concentrations for future comparison testing (eg, samples with concentrations that are infrequently encountered). Preexamination variables and appropriate storage conditions must be taken into consideration when storing samples. Caution should be exercised when storing frozen samples. Freezers with automatic defrost cycles should not be used, because these operate by periodically warming the compartment, partially thawing the contents, then refreezing. Samples should be stored below −70°C[21] to ensure immobilization of solid water and suitable stability of the frozen condition.

Use care when transporting samples between measurement systems to ensure appropriate stability of the analyte of interest and to prevent evaporation. When there is a transportation delay between locations that need to measure the same sample, it is recommended to prepare aliquots and coordinate the testing so all aliquots are measured at approximately the same time. It is generally not recommended to base an evaluation on a value measured at a significantly different time from other measurements. Determination of what constitutes a significantly different time depends on the stability characteristics of a given analyte.

7 Acceptance Criteria for Comparability Testing of Patient Results

There are no universally accepted criteria for evaluating the results of comparability testing; therefore, the laboratory director must determine the limits of acceptable differences for results produced by different measurement systems for the same analyte. The choice of criteria may vary from analyte to analyte and heavily depends on the availability of published information for each analyte (eg, clinical studies, biological variability data, and data from external PT/EQA programs). The primary objective is agreement between results from different measurement systems that is acceptable for the clinical situations in which the results will be interpreted. However, the inherent performance characteristics of the measurement systems should be taken into consideration when establishing acceptance criteria. If system capability is insufficient to meet desired comparability criteria, the frequency of comparability monitoring event failures will be impracticably high. Under those circumstances, improvement of measurement system performance (ie, replacement of a measurement system or optimization of current system operation) would be required to meet the desired comparability criteria. Circumstances may also arise when criteria based on the imprecision of the measurement systems being compared may be more stringent than necessary for clinical requirements, and the laboratory director may choose to base the acceptance criteria on clinical requirements.

Analytical goals for comparability of results can be defined using clinical approaches, expert opinion, or statistical approaches. The goal for comparability may vary depending upon the clinical use of an assay. Greater agreement among results is required when a result is used to identify changes in an individual patient over time vs use of a result as a component of an initial diagnostic workup. Therefore, comparability testing of a measurement system across different laboratory instruments that may be used to monitor the same patient over time should use comparability criteria consistent with that required for serial patient monitoring. A consensus hierarchy of approaches to establish criteria for analytical performance and measurement system comparability has been proposed by the International Organization for Standardization (ISO) Technical Committee 212 Working Group on Analytical Performance Goals Based on Medical Needs and members of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC).[22] The following approaches are listed in order of preference.

7.1 Evaluation of Comparability Based on Clinical Outcomes

Acceptance criteria based on well-designed clinical outcome studies are the highest standard for evaluating comparability testing. The strength of this approach is the clinical impact of analytical performance, which cannot be ignored. The weakness of the approach is clinical outcome studies are difficult to perform; therefore, there are very few examples in laboratory medicine.[23-27]

Example 7.1

An example of an assay with clinical outcomes–based data that can be used to make comparability recommendations is the use of the hemoglobin A1c assay (HbA1c) for monitoring an individual's diabetes control. The Diabetes Control and Complications Trial on Clinical Outcomes Related to HbA1c indicated that an HbA1c of 8.0% has a poorer clinical outcome compared to an HbA1c of 7.0%, and should therefore be accompanied by a change in patient management. Total error between methods should be kept to below ± 1% (ie, an absolute change of 1% reporting unit) so a patient with a poor clinical outcome (HbA1c ≥ 8.0%) cannot be misclassified as a well-controlled diabetic (HbA1c ≤ 7.0%). As total error includes both the imprecision and bias, when both measurement systems have low imprecision, some bias between methods may be tolerated.

7.2 Evaluation of Comparability Based on Clinician's Questionnaire

An alternative approach is to survey clinicians by questionnaire in order to determine their expectations of analytical quality that would give them confidence in managing patients.[28-31] The goal is a clinical consensus of the magnitude of change in an individual patient's results that would result in a change in clinical management. The advantage of this approach is it is based on clinical experience and therefore acceptable to clinicians. The disadvantages of the approach are that clinicians' expectations may be based on prevailing standards of analytical comparability, and also that the method still requires a rigorous methodology to define the clinical scenario and its analytical correlate.

Example 7.2

An example would be a survey that indicated that clinicians interpreted a 10% change (eg, a change in HbA1c concentration from 8.0% to 7.2%) in the HbA1c result as a significant change in a patient's clinical condition.[32] A total error goal less than 10% would consist of the known imprecision of both methods as well as the acceptable bias between the methods.

7.3 Evaluation of Comparability Based on Biological Variability

When monitoring a patient using two methods for the same analyte, each method may have analytical imprecision that is within desirable limits (ie, $< 0.5\ CV_I$), but bias between the methods may significantly reduce comparability. Therefore, for two methods that individually achieve desirable analytical

imprecision as defined above, a quality specification for the allowable difference between two methods measuring the same analyte can be defined as: **allowable difference < 0.33 × CV_I.**[33]

Appendix D shows the derivation of the desirable limits for difference between two methods for the same analyte using the principles of biological variability. Values for CV_G and CV_I are available for most common analytes.[34,35] Detailed discussions of this approach are available in numerous publications.[33,36-41]

The strength of the biological variability approach is it uses a defined statistical approach, uses measurable biological variability parameters, and takes into account the impact of measurement error on clinical interpretation of results. The weaknesses of this approach include that it is not based on clinical outcomes, and the necessary parameters for the calculations are not available for all analytes. In addition, the analytical performance of currently available methods is insufficient to meet the goals for desirable bias defined by the biological variability approach for several common analytes (eg, serum/plasma sodium and calcium measurements).

Example 7.3

For the monitoring of HbA1c, where the within-subject biological variability (CV_I) is 5.6%, the allowable difference between two methods that have achieved desirable imprecision is: allowable difference < 0.33 × CV_I or allowable difference < 1.8% (eg, if the HbA1c measurement on a patient sample by one method is equal to 8.0%, HbA1c values for comparison methods would need to be between 8.1% and 7.9% to meet the requirement).

7.4 Evaluation of Analytical Performance Based on Published Professional Recommendations

National or international professional expert bodies make judgments regarding what is an acceptable bias. The guidelines produced by these bodies are based on the combined expert understanding of the profession, often expressed as a consensus finding.

The strength of this approach is its inclusion of the understanding of experienced and reputable experts of analytical differences that could result in differing clinical interpretation of results. The weakness of this approach is that it is neither statistically rigorous nor based on clinical outcome studies. Another possible weakness is that these bias goals may refer to the difference between a method and a traceable target value rather than the bias between two methods.

Example 7.4

The National Cholesterol Education Program states that acceptable bias between cholesterol methods is ≤3% at 200 and 240 mg/dL (5.2 and 6.24 mmol/L).

7.5 Evaluation of Analytical Performance Based on Goals Set by Accrediting Agencies

Accrediting and regulatory agencies may define acceptable goals for imprecision or inaccuracy that are based on a combination of observed performance from PT/EQA data and advice obtained from industry and professional leaders or advisors.

The strength of this approach is it takes into account both the capability observed in industry as a whole, as well as the informed advice from industry and professional advisors that include what should be achievable in all laboratories. The weakness of this approach is it reflects what can be achieved rather than what is clinically required. Another possible weakness is these bias goals may refer to the difference between a method and a traceable target value rather than the bias between two methods.

Example 7.5

The Royal College of Pathologists of Australasia and the Australian Association of Clinical Biochemists state that in the Australian Quality Assurance Program in Chemical Pathology, the allowable limits of performance for total cholesterol is ± 0.5 mmol/L up to 10 mmol/L and ± 5% above 10 mmol/L (± 19 mg/dL up to 387 mg/dL and ± 5% above 387 mg/dL).

7.6 Evaluation of Analytical Performance Based on the General Capability

In this approach, performance that is similar to that of peers is defined as acceptable. Biases between measurement systems that are within the usual range of differences observed for similar measurement systems are defined as acceptable, because the industry and profession already accept those differences existing between laboratories and measurement systems.

The strength of this approach is that the information is readily accessible from PT/EQA results. The approach's weakness is that large differences may often be seen in PT/EQA schemes and some may be due to matrix errors. There is also no assessment made of the possible differences in clinical interpretation that could result from the differences observed. Similarly, differences between laboratories may be corrected for by differences in reference intervals and decision limits that are not evident in results from PT/EQA schemes.

Example 7.6

PT/EQA testing shows that the bias from a target value for HbA1c is ≤ 0.13% (an absolute difference of 0.13% reporting units) for the best 20% of laboratories. Fifty percent of laboratories have a bias of ≤ 0.29% (an absolute difference of 0.29% reporting units) and 10% of laboratories have a bias of > 0.77% (an absolute difference of 0.77% reporting units) that may be considered less acceptable.

8 Statistical Evaluation of Comparability Data

Analysis of comparability data does not always require sophisticated statistical analysis. Inspection of a simple plot or table of comparison data should be the first step in any evaluation of comparability data. This may be sufficient to assure the medical director that assays are performing in a comparable manner. It is left to the laboratory director's discretion to determine when a more rigorous analysis is required.

8.1 Hypothesis Testing

This document's statistical procedure for assessing the comparability of laboratory methods employs hypothesis testing. *Hypothesis testing* involves stating a null hypothesis (usually that the laboratory methods produce equivalent results), calculating a statistic, and rejecting the null hypothesis if the value of the statistic is highly unlikely when the null hypothesis is true. The *significance level* of a hypothesis test is the probability of incorrectly rejecting the null hypothesis when it is actually true, also called *Type I error*. The significance level is usually selected before conducting a hypothesis test. *Power* is the probability of correctly rejecting the null hypothesis when it is false. Incorrectly accepting the null hypothesis when it is actually false is called *Type II error*. Power is a property of the hypothesis test design and is useful for understanding the reliability of a hypothesis test.

An assumption of the hypothesis testing procedure in this document is that for any given specimen and laboratory method, replicate results are characterized by a normal distribution with some mean and variance. In reality, the mean and variance are typically unknown, but can be estimated from replicate measurements. The SD is defined as the square root of the variance.

Alternatively, the SD can be estimated using data derived from long-term QC testing (see Section 8.2.2). Refer to Appendix C for a more detailed discussion of statistical concepts.

8.2 Statistical Analysis of Comparability Data

Traditional approaches to method comparison (eg, Student's *t*-test, linear regression) are not easily adaptable to the simultaneous statistical comparison of multiple instruments, and typically require large sample sizes. This document presents an intuitive, simple, and statistically valid approach for the collection and simultaneous analysis of method comparison data from multiple instruments. The method uses comparisons at preselected analyte concentrations through parallel, replicate testing of a single specimen on two or more instruments. All of the results produced by the instruments are then compared using the range test. This approach minimizes the impact on the laboratory of comparability testing, yet provides adequate detection of clinically significant differences between instruments. Comparisons are made at analyte concentrations where reliable estimates of measurement imprecision are available to the laboratory (see Sections 8.2.2 and 8.2.3).

8.2.1 The Range Test

The measurements most people think about when the field of statistics is mentioned are sample mean (or average) and SD. Such measurements are extremely powerful in characterizing a population, especially if the number of times this population is sampled is large. However, when the sample size is small, such measurements may not be the most efficient way to either characterize a population or to perform a hypothesis test.

A number of efficient statistical measurements called shortcut procedures[42] based on order statistics are available for instances in which only a small sample size is available. One such measurement is the range. In the case of comparing two or more measurement systems using multiple samples (replicates) there are nk observations X_{ij} ($i = 1,\ldots, n$; $j = 1,\ldots, k$) where n is the number of measurement systems and k is the number of matched samples on each system. The range can be computed by first taking the mean $\overline{X}_i$ for each system over the k matched samples. The range w is thus:

$$w = \overline{X}_{\max} - \overline{X}_{\min}$$

where $\overline{X}_{\max}$ is the maximum observed mean and $\overline{X}_{\min}$ is the minimum observed mean.

The null hypothesis for the range test is that the true measurement system means μ_i are all equal. The alternative hypothesis is that the range of the true measurement system means is greater than or equal to the critical difference (CD).

All potential configurations of n measurement systems can be described mathematically and the Type II error from a range test can thus be computed by integrating across all possible configurations of means of the measurement systems. However, a far more efficient, and practical, way to determine Type II error and thus sample size is to propose a configuration on which to perform these calculations.

The least favorable configuration[43] for a range equal to the CD (ie, that configuration resulting in the greatest potential Type II error) has a central set of measurement systems where all means are equal and two measurement systems with means equally far apart in opposite directions from this central set of means ($\mu - CD/2$, μ, μ, …, μ, μ, $\mu + CD/2$). The most favorable is where half of the system means are equal to $\mu - CD/2$ and half are equal to $\mu + CD/2$. A far more likely scenario (1 method out scenario) has a set of measurement systems where all but one system has the same mean and the remaining measurement system has a mean that is removed from this common mean; ($\mu - CD$, μ, …, μ) or (μ, …, μ, $\mu + CD$).

Statistical tables (see Appendix B) have been provided to assist the laboratory in designing a comparison experiment by determining the numbers of runs and replicates to be run to attain a power of 80% and detect a defined difference between instrument results at an alpha level of 0.05. Knowledge of the within-run SD (SR) and total SD (ST) are required. Estimates of imprecision should be derived from a validation comparable to that proposed in CLSI document EP05,[44] performed either by the measurement system manufacturer or the laboratory. However, ST may be estimated from long-term CVs (≥ six months) derived from control material data, and SR may be estimated from at least a 20-replicate run of the appropriate material.

8.2.2 Estimating Variance Over the Analytical Measurement Range Using Quality Control Data

Data from statistical QC monitoring can be used to estimate the variance, the SD, or the CV at concentrations or activities near those of the QC materials. Data should be collected for at least six months to ensure most sources of variability in the measurement procedure are represented in the imprecision estimate. Sources of variability include calibration events; changes in lot of calibrator material; reagent changes during use; changes in lot of reagent; changes in components such as pipettes, temperature control, washing systems, and detection devices; maintenance cycles; environmental factors such as temperature and humidity; and fluctuations in electrical power.

The QC materials used to estimate variance may deteriorate during their open bottle use period and during long-term storage. If these conditions occur, the estimate of the SD or CV applicable to a patient sample may be artifactually increased. Despite this potential limitation, QC data are generally the best available sources for estimating the variance (imprecision) of a method. Data from PT/EQA suggest that estimates of imprecision derived from commercially prepared control materials are comparable to estimates derived from frozen serum.[45]

8.2.2.1 Estimation of a Measurement System Precision Profile

Available QC materials often do not challenge the full AMR of a measurement system. When variance estimates are needed for concentrations or activities that differ from those of the QC materials, additional procedures are necessary. Consultation with measurement system manufacturers about available imprecision estimates at other concentrations or activities may be helpful. Laboratories may also produce estimates of imprecision using pools of patient samples, calibrator, linearity, or PT/EQA materials following CLSI document EP05[44] (20-day protocol), at one or more concentrations or activities in the extended range likely to be encountered in samples to be used for verification of comparability between methods. However, a 20-day protocol is unlikely to account for a number of the variability sources described above, and cumulative data over a longer time period (eg, six months) are preferred. CLSI document EP15[46] (five-day protocol) is not adequate in this situation, because that procedure is intended primarily to verify manufacturers' claims, not to estimate the imprecision as a basis for acceptance criteria for another statistical test.

It is recommended to determine the SD at an adequate number of concentrations or activities over the range that will be used for comparison evaluation among measurement systems. In many cases, the SD may be approximately constant over a range of concentrations or activities, or may be proportional to the concentration or activity, in which case the CV will be reasonably constant. In either case, development of a table or graph of the SD at different concentrations or activities to use in the range test statistic is suggested. When there is a consistent change in SD with concentration or activity, interpolation between the values determined will allow good estimates of SD at intermediate concentrations or activities. Particular caution needs to be paid to low concentrations or activities where the SD may increase substantially as the concentration of activity approaches the lower limit of the measurement range. An "imprecision profile" for an analyte measured on a given measurement system needs to be established only once. It is reasonable to assume the variance will be approximately the same at subsequent time periods as long as QC indicates the measurement system continues to meet its specifications.

For concentrations or activities that approach the lower limit of the AMR, it may be necessary to base acceptance criteria on an absolute difference in concentration or activity units between results for two or more measurement systems. The absolute difference may be based on medical usefulness criteria rather than statistical performance criteria. The laboratory director will need to establish the medical usefulness criteria based on the population served by the laboratory.

8.2.3 Identifying Concentrations Suitable for Use in the Comparison Evaluation

For the range test comparison protocol, patient samples are selected with concentrations or activities that are within the interval of the "precision profile" for an analyte measured on a given measurement system (see Section 8.2.2). If a "precision profile" has not been established, the laboratory should select patient samples with concentrations or activities that are "reasonably close" to the nominal values for QC materials, or other values for which the SD is known. "Reasonably close" is difficult to specify but, in general for values reported to at least two significant digits, values within 20% will likely have similar SD to that of the QC or other material. A "precision profile" is recommended for samples reported to a single digit or that approach the lower limit of the AMR.

A special procedure is needed when different measurement systems are to be compared and they use QC materials that have different nominal analyte values and, thus, different SDs at those values. In this situation, it is recommended to prepare pools of patient samples and measure aliquots from each pool according to CLSI document EP05[44] to establish an SD value for each measurement system at the nominal concentration of the pool.

8.3 Fixed Limit Evaluation

In some situations, it may be desirable to establish a fixed limit for the agreement between the largest and smallest numeric values observed among a group of measurement systems being compared. Analyte concentrations that approach the lower limit of quantitation are a common situation in which a criterion based on a percent difference is not suitable, because the small magnitude of the numeric result causes a small absolute difference to be a large percent difference. In such situations, the concepts described in Section 7 should be considered, but the fixed limit generally requires judgment of the clinical impact of differences at numeric values for which a percent criterion is not realistically achievable.

For example, CLSI document C30[47] states: "Ideally, 95% of the individual results from the POC glucose monitoring system should agree within ± 15 mg/dL (± 0.83 mmol/L) of the laboratory analyzer values at glucose concentrations below 75 mg/dL (4.2 mmol/L) and within ± 20% of the laboratory analyzer values at glucose concentrations at or above 75 mg/dL (4.2 mmol/L)." In this example, the ± 15 mg/dL (± 0.83 mmol/L) fixed criterion below 75 mg/dL (4.2 mmol/L) reflects the state of the art in point-of-care testing (POCT) glucose devices at the time the guideline was published.

9 Point-of-Care Testing

POCT presents unique challenges for comparability testing. Large numbers of instruments may have to be compared against a laboratory instrument (reference instrument), and reagents are frequently expensive. Methods to be compared often analyze different sample types (eg, whole blood vs plasma or serum), require different sample acquisition techniques (eg, venipuncture vs fingerstick), or employ nontraditional analysis (eg, measurement of transcutaneous bilirubin). On the other hand, laboratory testing and POCT are frequently performed on specimens collected in close temporal proximity to each other (eg, result confirmation testing), providing an opportunity for comparison.

9.1 Specimen Selection

Whenever possible, comparisons should be performed with simultaneously obtained specimens of the proper type for the measurement systems involved (eg, fingerstick [capillary] whole blood for a POCT glucose measurement system vs venous plasma measurement system), unless both systems use the same sample type. Otherwise, artifactual differences may be detected (due to a sample being inappropriate for one of the measurement systems), or differences that should be evident (due to the difference in sample sources normally used in the measurement systems) may be masked. For example, venous specimens may not be appropriate for analysis with POC glucose testing devices that employ a methodology that is dependent upon a minimum sample oxygen content. Conversely, other POC glucose testing methodologies may be suppressed by elevated oxygen concentrations and would not be appropriate for the analysis of arterial specimens.

9.2 Specimen Acquisition

POC comparison testing can often be accomplished by performing POCT analysis at the same time venous (or arterial) specimens are collected for laboratory analysis via phlebotomy. This approach has two major advantages:

1) Reduction of POCT performance variability by limiting the number of testing personnel

2) Elimination of the possibility for a change in the patient's condition (eg, glucose or insulin administration in the case of POC glucose testing) between the collection of the POCT and laboratory specimens

Disadvantages include:

1) Additional discomfort for the patient (if a capillary specimen is required for POCT)
2) More complex planning and coordination to execute
3) More difficulty in assessing the full range of test results
4) Failure to assess comparability across all testing personnel

Laboratories should follow local policies with respect to requirements for institutional review board approval and patient consent for the collection of additional specimens for method validation or comparison testing.

A variation of this approach that avoids additional sample collection (and possible patient discomfort), involves comparison testing performed on specimens that have been collected in vacuum tubes at the bedside and transported to the laboratory for analysis. This approach is only applicable if the manufacturer has validated testing of anticoagulated, noncapillary specimens on POCT instruments. This approach overcomes the disadvantages noted above, except it still fails to assess comparability across all testing personnel. In addition, POCT measurement of an analyte in venous (or arterial) anticoagulated whole blood may not mimic measurements made on capillary specimens, and influences of a capillary specimen on analyte values is not evaluated. Note that safety concerns generally limit testing of vacuum tube specimens to the laboratory, where tubes can be safely decapped behind a protective screen to acquire an aliquot of blood for POCT.

Another approach monitors the agreement between results from POCT at the bedside with near simultaneously collected blood sent to the laboratory for analysis. Laboratory analysis of blood collected proximal in time to a POCT measurement generally gives acceptable agreement, provided the laboratory-analyzed specimens have been properly collected and handled to ensure analyte stability before analysis. How close in time the two specimens must be collected depends upon the analyte in question and the

probability of a change in analyte concentration between the samplings. Advantages of this approach include:

1) Relatively frequent assessment of comparability
2) Assessment of the entire testing process
3) Assessment of a greater range of test results
4) Assessment of all testing personnel on all shifts and days of testing
5) Assessment of all measurement systems

Disadvantages to this approach include:

1) Potential introduction of additional variability due to changes in the patient's condition between collection of the laboratory and POCT specimens (eg, insulin or glucose administration for glucose POCT)

2) Introduction of variability due to a lack of synchronization of clocks used to identify sample collection times

3) Inability to differentiate noncomparability due to operator error from that due to analytical error

In addition, to evaluate the comparability of individual POCT measurement systems, results must be tracked by individual devices in use. Confirmation of potential noncomparability using this approach should be confirmed using one of the simultaneous testing protocols described previously.

9.3 Range of Specimen Values

Test results ideally should span the AMR of the system in question. At a minimum, low and high concentrations (eg, upper and lower reference interval limits) or those around clinical decision points should be analyzed.

9.4 Multiple Devices of the Same Make and Model

For POCT programs with a large number of devices of the same make and model in use (eg, glucose meters), an alternative approach is acceptable for documenting comparability of patient results between the POCT devices and the laboratory. When all of the POCT measurement systems of the same make and model to be evaluated are using one lot of reagent strips/cartridges, comparability testing can be performed with a representative subset of POCT devices while simultaneously evaluating QC and/or PT/EQA results among all of the POCT devices. If comparability of patient results among the subset of POCT devices included is confirmed, the comparability of the other POCT devices of the same make and model can be inferred from acceptable agreement of QC and/or PT results for the same lots of reagent strips/cartridges and QC or PT materials. If more than one device type or lot of reagents is used, this process must be repeated for each combination. In subsequent comparability evaluation events, different subsets of instruments should be tested so, over time, all POCT devices will have comparability testing performed on them.

9.5 Statistical Considerations for Point-of-Care Comparability Testing

For circumstances in which the range test protocol is not applicable (eg, sample types required by laboratory and POCT measurement systems are different; simultaneous comparison of multiple specimens), alternative statistical approaches for comparison of measurement systems have been described.[48,49]

10 Range Test Comparability Protocol

This protocol is designed for studies comparing up to 10 measurement systems with a maximum of two distinct runs (as defined by the laboratory) and varying numbers of replicates based on system performance. If the criteria for the range test comparability protocol cannot be met, alternative approaches should be considered, such as the protocols described in CLSI document EP15[46] or CLSI document EP09.[1] The following protocol should be performed for at least two analyte concentrations on each measurement system per comparability testing event.

10.1 Select an Analyte for Comparison

Any analyte measured by more than one test system in a health care system should be considered for comparability testing.

10.2 Select the Instruments to Be Compared

Ideally, all measurement systems that are currently in use for measuring patient samples for the analyte should be compared.

10.3 Identify an Approximate Analyte Concentration for Comparison Testing

1) Select an analyte concentration for which estimates of imprecision are known for each of the measurement systems. Estimates of SR and ST are required. Estimates of imprecision should be derived from a validation comparable to that proposed in CLSI document EP05,[44] performed either by the measurement system manufacturer or the laboratory. However, ST may also be estimated from long-term CVs (≥ six months) derived from QC data, and SR may be estimated from at least a 20-replicate run of the appropriate material (see Section 6). Record the ST and SR values of the assays to be tested.

2) An underlying assumption is that a single estimate of measurement system variability can be used to characterize all of the systems being compared. To test this assumption, compare the magnitudes of the ST values of the measurement systems to determine if the greatest and smallest ST values differ by less than twofold (ie, 2×). If so, proceed to the next step. If the ST values differ by greater than twofold, consider whether the system with the largest ST value must be included in the comparison or whether a different, perhaps higher, analyte concentration would provide more comparable system ST values. If neither change is feasible, be aware that sample size requirements and the potential for Type I error may be increased. Alternatively, consider protocols from CLSI document EP15[46] or CLSI document EP09[1] for demonstration of measurement system comparability.

3) Calculate pooled ST and SR values from the measurement system SDs:

   Pooled $ST = ([ST_1^2 + ST_2^2 + \ldots ST_n^2]/n)^{1/2}$, where there are n measurement systems.

   Pooled $SR = ([SR_1^2 + SR_2^2 + \ldots SR_n^2]/n)^{1/2}$, where there are n measurement systems.

NOTE: Calculation of a pooled ST using this equation assumes the long-term ST values for the measurement systems (ST_1, ST_2, ….ST_n) are all calculated from approximately equal numbers (ie, equal sample sizes).

10.4 Calculate the Desired Concentration or Activity to Be Used for Comparison Sample Selection

Determine the mean concentration or activity value of the control (or other) material that will be used to estimate imprecision for each of the measurement systems. Calculate the grand mean of the measurement system mean values for the material and record it. Use the grand mean as the comparison sample desired concentration or activity for Section 10.5:

- Comparison sample desired value = grand mean of control material means = (mean control material concentration for Analyzer A + mean control material concentration for Analyzer B + … + mean control material concentration for Analyzer J) / number of analyzers to be compared.

If a precision profile has been determined, the desired concentration is within the range of values over which the SD is approximately constant (see Section 8.2.2.1).

10.5 Select a Sample for Comparison Testing

Identify a specimen that: 1) meets the stability requirements of the analyte for all assays; 2) does not contain substances that interfere with the assays being compared; 3) has sufficient volume for testing; and 4) has an estimated value (based on testing on any one of the measurement systems to be compared) within 20% of the test sample target value calculated in Section 10.4 (see Section 8.2.3 for exceptions to this approach). If a large number of measurement systems are to be compared, pooled samples may be used (see Section 6.1.1 for potential limitations of pooled specimens). If measurement systems to be evaluated are located remotely from each other, be sure to stabilize the specimen appropriately for transportation.

10.6 Select the Appropriate Level of Acceptance Criteria That Can Be Applied to the Comparison Test (From Section 7)

1) Determine if there are recommendations based on clinical outcome studies that are within the performance specifications of the measurement systems being compared (ie, the ST values of the assays to be compared are less than the recommended acceptance criteria); if not, proceed to the next level of evidence.

2) Determine if the clinicians at the institution(s) have specific recommendations based on their clinical experience that are within the performance specifications of the methods being compared; if not, proceed to the next level of evidence.

3) Determine if there are recommendations based on biological variability that are within the performance specifications of the methods being compared; if not, proceed to the next level of evidence.

4) Determine if there are minimal requirements set by an accreditation agency; if not, proceed to the next level of evidence.

5) Determine the analytical capability of the measurement system based on external PT (EQA) data; if no data are available, proceed to the next level of evidence.

6) If no external comparability criteria are applicable, determine the analytical capability of the measurement system based on internal imprecision data.

10.7 Calculate the Critical Difference for the Comparability Test

To calculate the CD, multiply the estimated value of the sample selected in Section 10.5 by the acceptance criterion (as a decimal rather than percent) selected in Section 10.6.

10.8 Determine the Number of Runs and Replicates to Be Run and the Range Rejection Limit

1) Calculate the following ratios: CD/ST and SR/ST.

2) Consult the appropriate table in Appendix B for the number of measurement systems to be compared. Determine the number of runs to be performed and the number of replicates per run. Calculate the range rejection limit by multiplying the corresponding coefficient listed in the table by the CD. See Appendix B for an explanation of the tables. **NOTE:** The power listed under the most likely scenario (1 method out) is recommended, but for a more conservative set-up calculation, the power listed under the least favorable configuration can be used.

10.9 Perform the Comparison

1) Analyze the specimen selected in Section 10.5 on each of the measurement systems to be compared, performing the number of runs and replicates specified in Section 10.8.

2) If replicate analyses are not indicated, the individual results from each measurement system will be compared directly.

3) If replicate analyses are performed, calculate the mean value from the replicate analyses of the specimen separately for each measurement system.

4) Calculate the range as the difference between the most disparate measurement system means or individual values if replicate measurements are not indicated:

   - When replicate measurements are made: highest measurement system mean − lowest measurement system mean

   - When replicate measurements are not made: highest measurement system value − lowest measurement system value

5) Compare the calculated range with the range rejection limit determined in Section 10.8.

6) If the calculated range is less than or equal to the range rejection limit, conclude that all measurement systems perform comparably at the analyte level evaluated.

7) If the calculated range is greater than the range rejection limit, conclude that the measurement systems with the most disparate mean values (or individual values) perform significantly different from each other. Compute the median of all the systems' values. Compute the absolute difference of each system's value from this median value. In troubleshooting, the demonstrated noncomparability (see Section 10.11) concentrates on the systems with the highest such absolute difference.

8) If a measurement system is known to produce results with an expected bias vs another measurement system, calculate the range as noted below and compare the resulting value with the CD to determine if the known and expected bias is greater than expected (see Example 4 in Appendix A):

- When replicate measurements are made: (highest measurement system mean − lowest measurement system mean) − expected absolute difference

- When replicate measurements are not made: (highest measurement system value − lowest measurement system value) − expected absolute difference

10.10 Evaluate the Clinical Relevance of the Comparison Results

The medical director must assess the medical significance of any statistically significant comparison differences (see Section 8 and Appendix C for a discussion of Type I and Type II errors).

10.11 Troubleshooting Noncomparability

If a measurement system is determined to be noncomparable, troubleshoot any analytical problems and repeat the comparison.

References

1 CLSI. *Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition (Interim Revision)*. CLSI document EP09-A2-IR. Wayne, PA: Clinical and Laboratory Standards Institute; 2010.

2 CLSI. *Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions; Approved Guideline—Third Edition*. CLSI document C24-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2006.

3 CLSI. *Metrological Traceability and Its Implementation; A Report.* CLSI document X05-R. Wayne, PA: Clinical and Laboratory Standards Institute; 2006.

4 ISO. *In vitro diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials*. ISO 17511. Geneva, Switzerland: International Organization for Standardization; 2003.

5 Miller MJ, Astles JR, Baszler T, et al.; National Center for Emerging and Zoonotic Infectious Diseases, CDC. Guidelines for safe work practices in human and animal medical diagnostic laboratories. *MMWR Surveill Summ.* 2012;61 Suppl:1-102.

6 CLSI. *Protection of Laboratory Workers From Occupationally Acquired Infections; Approved Guideline—Third Edition*. CLSI document M29-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2005.

7 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 August 15, 2012.

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

9 ISO. *Statistics – Vocabulary and symbols – Part 1: Probability and general statistical terms and terms ued in probability*. ISO 3534-1. Geneva, Switzerland: International Organization for Standardization; 2006.

10 ISO. *In vitro diagnostic medical devices – Measurement of quantities in samples of biological.origin – Requirements for certified reference materials and the content of supporting documentation.* ISO 15194. Geneva, Switzerland: International Organization for Standardization; 2009.

11 ISO. *Quality management systems – Fundamentals and vocabulary*. ISO 9000. Geneva, Switzerland: International Organization for Standardization; 2005.

12 ISO. *Medical laboratories – Requirements for safety.* ISO 15190. Geneva, Switzerland: International Organization for Standardization; 2003.

13 ISO/IEC. *Safety aspects – Guidelines for their inclusion in standards.* ISO/IEC Guide 51. Geneva, Switzerland: International Organization of Standardization; 1999.

14 ISO. *Medical laboratories – Particular requirements for quality and competence*. ISO 15189. Geneva, Switzerland: International Organization for Standardization; 2007.

15 Linnet K. The exponentially weighted moving average (EWMA) rule compared with traditionally used quality control rules. *Clin Chem Lab Med*. 2006;44(4):396-399.

16 Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. *Clin Chim Acta*. 2003;327(1-2):25-37.

17 Miller WG, Myers GL, Rej R. Why commutability matters. *Clin Chem*. 2006;52(4):553-554.

18 CLSI/NCCLS. *Preparation and Validation of Commutable Frozen Human Serum Pools as Secondary Reference Materials for Cholesterol Measurement Procedures; Approved Guideline*. CLSI/NCCLS document C37-A. Wayne, PA: NCCLS; 1999.

19 Lawson NS, Williams TL, Long T. Matrix effects and accuracy assessment. Identifying matrix-sensitive methods from real-time proficiency testing data. *Arch Pathol Lab Med*. 1993;117(4):401-411.

20 Long T. Statistical power in the detection of matrix effects. *Arch Pathol Lab Med*. 1993;117(4):387-392.

21 Jennings TA. Effect of formulation on lyophilization, part 2.http://www.ivdtechnology.com/article/effect-formulation-lyophilization-part-2. Accessed August 15, 2012.

22 Fraser CG, Kallner A, Kenny D, Petersen Ph. Introduction: strategies to set global quality specifications in laboratory medicine. *Scand J Clin Lab Invest.* 1999;59(7):477-478.

23 Larsen ML, Fraser CG, Petersen PH. A comparison of analytical goals for haemoglobin A1c assays derived using different strategies. *Ann Clin Biochem*. 1991;28(Pt 3):272-278.

24 Jenny RW. Analytical goals for determination of theophylline concentration in serum. *Clin Chem*. 1991;37(2):154-158.

[25] Petersen PH, Hørder M. Ways of assessing quality goals for diagnostic tests in clinical situations. *Arch Pathol Lab Med.* 1988;112(4):435-443.

[26] Hyltoft Petersen P, Hørder M. Influence of analytical quality on test results. *Scand J Clin Lab Invest.* 1992;52 Suppl 208:65-87.

[27] Nordic Clinical Chemistry. Medical need for quality specifications in clinical laboratories: report of a NORDKEM (Nordic Clinical Chemistry) project. *Ups J Med Sci*. 1993;98(3):189-491.

[28] Barnett RN. Medical significance of laboratory results. *Am J Clin Pathol*. 1968;50(6):671-676.

[29] Elion-Gerritzen WE. Analytical precision in clinical chemistry and medical decisions. *Am J Clin Pathol.* 1980;73(2):183-195.

[30] Skendzel LP, Barnett RN, Platt R. Medically useful criteria for analytic performance of laboratory tests. *Am J Clin Pathol*. 1985;83(2):200-205.

[31] Thue G, Sandberg S, Fugelli P. Clinical assessment of haemoglobin values by general practitioners related to analytical and biological variation. *Scand J Clin Lab Invest.* 1991;51(5):453-459.

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

[33] Petersen PH, Fraser CG, Westgard JO, Larsen ML. Analytical goal-setting for monitoring patients when two analytical methods are used. *Clin Chem*. 1992;38(11):2256-2260.

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

[35] Westgard QC. Desirable specifications for total error, imprecision, and bias, derived from intra- and inter-individual biologic variation. http://www.westgard.com/biodatabase1.htm. Accessed August 15, 2012.

[36] Cotlove E, Harris EK, Williams GZ. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. 3. Physiological and medical implications. *Clin Chem*. 1970;16(12):1028-1032.

[37] Fraser CG, Petersen PH. Desirable standards for laboratory tests if they are to fulfill medical needs. *Clin Chem.* 1993;39(7):1447-1455.

[38] Fraser CG, Petersen PH, Ricos C, Haeckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Clin Biochem*. 1992;30(5):311-317.

[39] Klee GG. Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay specificity. *Clin Chem*. 1993;39(7):1514-1518.

[40] Stöckl D, Baadenhuijsen H, Fraser CG, Libeer JC, Petersen PH, Ricós C. Desirable routine analytical goals for quantities assayed in serum. Discussion paper from the members of the external quality assessment (EQA) Working Group A on analytical goals in laboratory medicine. *Eur J Clin Chem Clin Biochem*. 1995;33(3):157-169.

[41] Ricós C, Iglesias N, García-Lario JV, et al. Within-subject biological variation in disease: collated data and clinical consequences. *Ann Clin Biochem.* 2007;44(Pt 4):343-352.

[42] David HA, Nagaraja HN. *Order Statistics.* 3rd ed. Hoboken, NJ: Wiley InterScience; 2003.

[43] Chen SY, Chen HJ. A range test for the equivalency of means under unequal variances. *Technometrics.* 1999;41(3):250-260.

[44] CLSI/NCCLS. *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition*. CLSI/NCCLS document EP05-A2. Wayne, PA: NCCLS; 2004.

[45] Ross JW, Miller WG, Myers GL, Praestgaard J. The accuracy of laboratory measurements in clinical chemistry: a study of 11 routine analytes in the College of American Pathologists Chemistry Survey with fresh frozen serum, definitive methods, and reference methods. *Arch Pathol Lab Med.* 1998;122(7):587-608.

[46] CLSI. *User Verification of Performance for Precision and Trueness; Approved Guideline—Second Edition*. CLSI document EP15-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2006.

[47] CLSI/NCCLS. *Point-of-Care Blood Glucose Testing in Acute and Chronic Care Facilities; Approved Guideline—Second Edition*. CLSI/NCCLS document C30-A2. Wayne, PA: NCCLS; 2002.

[48] Stöckl D, Dewitte K, Fierens C, Thienpont LM. Evaluating clinical accuracy of systems for self-monitoring of blood glucose by error grid analysis: comment on constructing the "upper A-line." *Diabetes Care*. 2000;23(11):1711-1712.

[49] Parkes JL, Slatin SL, Pardo S, Ginsberg BH. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care*. 2000;23(8):1143-1148.

Appendix A. Worked Examples

Example 1

A laboratory director wanted to evaluate the comparability of AST (aspartate transaminase) measurements between two analyzers of the same make and model.

- *Precision Estimates*

Precision estimates were derived from the manufacturer's validation data, as follows:

	Precision Estimates		
Analyzer	Control Mean, *U/L*	SR, *U/L*	ST, *U/L*
A	40.37	1.89	2.10
B	41.52	1.99	2.21
	Mean = 40.95	**Pooled SR = 1.94**[*]	**Pooled ST = 2.16**[†]

Abbreviations: SR, within-run standard deviation; ST, total standard deviation.
[*] Pooled within-run SD: $SR = ([\{1.89\}^2 + \{1.99\}^2]/2)^{1/2} = \sqrt{3.77} = 1.94$ *U/L*.
[†] Pooled total SD: $ST = ([\{2.10\}^2 + \{2.21\}^2]/2)^{1/2} = \sqrt{4.65} = 2.16$ *U/L*.

Because the respective ST values of the two analyzers differed by less than a factor of two, the range test protocol was applicable without further considerations.

- *Sample Selection*

The sample range was calculated as ±20% of the respective grand mean of the controls, or 32.76 to 49.14 U/L. A sample with an initial value of 37.80 on the Analyzer A was selected for the comparison test.

- *Acceptability Criteria*

To determine acceptability criteria following the hierarchy in Section 7 of this document, the laboratory director noted that recommendations based on clinical outcome studies do not exist, but key clinicians in the institution recommended comparability at a level of 10%.

- *Number of Runs, Replicates, and Range Rejection Limit*

To calculate the critical difference (CD), one multiplies the acceptability criterion by the initial value from Analyzer A: $CD = 37.80 \times 0.1 = 3.78$. The following ratios are then calculated: CD/Pooled ST = $3.78/2.16 = 1.75$ and $SR/ST = 1.94/2.16 = 0.90$. Consulting Table B1 (in Appendix B) entries for CD/ST of 1.75 and SR/ST of 0.90, one has a choice of one run with 160 replicates or two runs of five replicates per run. The medical director elected to collect data from two distinct runs (as defined by the laboratory). The range rejection limit (see Table B1) is calculated as $0.67 \times CD = 0.67 \times 3.78 = 2.53$.

Appendix A. (Continued)

- *Comparison Data*

The sample was analyzed with the following results:

Run	Replicate	Analyzer A	Analyzer B
1	1	37.8	40.4
	2	38.5	38.6
	3	38.0	38.8
	4	38.2	39.1
	5	38.4	40.2
2	1	38.6	38.6
	2	38.5	40.1
	3	38.0	39.5
	4	38.7	39.0
	5	38.5	38.0
Mean		**38.32**	**39.23**

Statistic	
Range	39.23 − 38.32 = **0.91**
Range Rejection Limit	**2.53**
Status	**Pass**

- *Conclusion*

Because the observed range was less than the range rejection limit, the laboratory director concluded that the comparability of the methods is acceptable.

Example 2

A laboratory director wanted to evaluate the comparability of total bilirubin measurements between two analyzers of the same make and model.

- *Precision Estimates*

Precision estimates were derived from the manufacturer's validation data, as follows:

	Precision Estimates		
Analyzer	Control Mean, mg/dL	SR, mg/dL	ST, mg/dL
A	0.95	0.03	0.03
B	1.05	0.03	0.03
	Mean = 1.00	**Pooled SR = 0.03**	**Pooled ST = 0.03**

Abbreviations: SR, within-run standard deviation; ST, total standard deviation.

Because the respective ST values of the three analyzers differed by less than a factor of two, the range test protocol was applicable without further considerations.

Appendix A. (Continued)

- *Sample Selection*

The sample range was calculated as ± 20% of the respective grand mean of the controls, or 0.08–1.20 mg/dL. A sample with an initial value of 0.98 mg/dL on Analyzer A was selected for the comparison test.

- *Acceptability Criteria*

In the absence of acceptability criteria based on clinical outcomes or clinician consensus, the laboratory director identified a criterion based on biological variability that specifies an allowable difference of 8.5% ($CV_I/3 = 25.6\%/3 = 8.5\%$).

- *Number of Runs, Replicates, and Range Rejection Limit*

To calculate the CD, one multiplies the acceptability criterion by the initial value from Analyzer A: CD = 0.98 × 0.085 = 0.08. The following ratios are then calculated: CD/ Pooled ST = 0.08/0.03 = 2.67 and SR/ST = 0.03/0.03 = 1.00. Consulting Table B1 entries for CD/ST of 2.67 and SR/ST of 1.00, the laboratory director chooses one run, but must use interpolation of the CD/ST values in Table B1 to determine the number of replicates and the range rejection limit. The number of replicates needed is given by the following equation:

$$\frac{1}{n} = \frac{1}{n_0} + (x - x_0)\frac{n_0 - n_1}{n_0 n_1 (x_1 - x_0)}$$

where $x = 2.67$ and n is unknown, $x_0 = 2.50$ and $n_0 = 3$, and $x_1 = 3.00$ and $n_1 = 2$.

Therefore: $1/n = 1/3 + (2.67 - 2.50)\ [\{3 - 2\}/\{3 \bullet 2(3.00 - 2.50)\}] = 0.386$; $n = 1/0.386 = 2.6$, or ≈ 3 replicates.

The coefficient for the range rejection limit calculation is derived as follows:

$$\frac{1}{L} = \frac{1}{L_0} + (x - x_0)\frac{L_0 - L_1}{L_0 L_1 (x_1 - x_0)}$$

where $x = 2.67$ and L is unknown, $x_0 = 2.50$ and $L_0 = 0.65$, and $x_1 = 3.00$ and $L_1 = 0.66$.

Therefore: $1/L = 1/0.65 + (2.67 - 2.50)\ [\{0.65 - 0.66\}/0.65 \bullet 0.66\,(3.00\text{-}2.00)\}] = 1.53$; $L = 1/1.53 = 0.65$

The corresponding range rejection limit is calculated as: $0.65 \bullet CD = 0.65 \bullet 0.08 = 0.05$.

- *Comparison Data*

The sample was analyzed with the following results:

Replicate	Analyzer A	Analyzer B
1	0.98	1.00
2	0.95	1.03
3	0.99	0.99
Mean	**0.97**	**1.01**

Appendix A. (Continued)

Statistic	
Range	1.01 − 0.97 = **0.04**
Range Rejection Limit	**0.05**
Status	**Pass**

- *Conclusion*

Because the observed range was less than the range rejection limit, the laboratory director concluded that the comparability of the methods is acceptable.

Example 3

A laboratory director wanted to evaluate the comparability of WBC measurements among three analyzers.

- *Precision Estimates*

Precision estimates were derived from long-term QC statistics (ST) and a QC lot validation (SR, n = 20) as follows:

Analyzer	Control Mean, $\times 10^3/\mu L$	SR, $\times 10^3/\mu L$	ST, $\times 10^3/\mu L$
A	3.53	0.068	0.080
B	2.61	0.054	0.063
C	2.88	0.085	0.100
	Mean = 3.01	**Pooled SR = 0.070**[*]	**Pooled ST = 0.082**[†]

Abbreviations: SR, within-run standard deviation; ST, total standard deviation.
[*] Pooled within-run SD: $SR = ([\{0.068\}^2 + \{0.054\}^2 + \{0.085\}^2]/3)^{1/2} = \sqrt{0.005} = 0.070$.
[†] Pooled total SD: $ST = ([\{0.080\}^2 + \{0.063\}^2 + \{0.100\}]/3)^{1/2} = \sqrt{0.007} = 0.082$.

NOTE: Analyzer A is the laboratory's main analyzer. Due to differences in capabilities, Analyzers B and C require a different set of controls than Analyzer A, so the control means are not directly comparable.

Because the respective ST values of the three analyzers differed by less than a factor of two, the range test protocol was applicable without further considerations.

- *Sample Selection*

The sample range was calculated as ± 20% of the respective grand mean of the controls, or $2.4 - 3.6 \times 10^3/\mu L$. A sample with an initial value of 3.40 on Analyzer A was selected for the comparison test.

- *Acceptability Criteria*

In the absence of acceptability criteria based on clinical outcomes or clinician consensus, the laboratory director identified a criterion based on biological variability that specifies an allowable difference of 3.63% ($CV_I/3 = 10.9\%/3 = 3.63\%$). This was considered impractical, as two runs of 27 replicates each would be required for the comparison test. The laboratory director determined that the only practical level of acceptance criteria available was based on goals set by regulatory authorities. In the United States, Clinical Laboratory Improvement Amendment regulations set a PT performance goal of 15% of the target value, while the German Medical Association maximal permissible deviation is 13%. The laboratory director selected 15% as the comparison criterion.

Appendix A. (Continued)

- *Number of Runs, Replicates, and Range Rejection Limit*

To estimate the CD, one multiplies the acceptability criterion by the initial value from Analyzer A: CD = 3.40 × 0.15 = 0.51. The following ratios are then calculated: CD/Pooled ST = 0.51/.082 = 6.22 and SR/ST = 0.070/.082 = 0.85. Consulting Table B2, a value for CD/ST ≥ 4.00 requires one run and one replicate. The range rejection limit is calculated as 0.83 × CD = 0.83 × 0.51 = 0.42.

- *Comparison Data*

The samples were analyzed with the following results:

Replicate	Analyzer A	Analyzer B	Analyzer C
1	3.40	2.90	3.22

Statistic	
Range	3.40 – 2.90 = **0.50**
Range Rejection Limit	**0.42**
Status	**Fail**

- *Conclusion*

Because the observed range was greater than the range rejection limit, the laboratory director concluded that the comparability of the methods is not acceptable. The laboratory director should institute an evaluation of Analyzer B, the most disparate measurement system, and take appropriate corrective action.

Example 4

A laboratory director wanted to evaluate the comparability of prothrombin time (PT) results between a main analyzer and backup analyzer. Previous data analyses showed that at an approximate PT value of 13 seconds, the backup analyzer (B) produces results that are about 10% longer than the main analyzer.

- *Precision Estimates*

Precision estimates were derived from long-term QC statistics (ST) and a QC lot validation (SR, n = 20) as follows:

Analyzer	Precision Estimates		
	Control Mean, *U/L*	SR, *U/L*	ST, *U/L*
A	13.5	0.31	0.32
B	14.9	0.33	0.34
	Mean = 14.2	**Pooled SR = 0.32***	**Pooled ST = 0.33†**

Abbreviations: SR, within-run standard deviation; ST, total standard deviation.
* Pooled within-run SD: $SR = ([\{0.31\}^2 + \{0.33\}^2]/2)^{1/2} = \sqrt{0.103} = 0.32$.
† Pooled total SD: $ST = ([\{0.32\}^2 + \{0.34\}^2]/2)^{1/2} = \sqrt{0.109} = 0.33$.

Appendix A. (Continued)

- *Sample Selection*

Sample ranges were calculated as 13 seconds ± 20%, or 10.4 to 15.6 seconds, which includes the grand mean of the controls. A sample with an initial value of 14.2 on the main analyzer was selected for the comparison test.

- *Acceptability Criteria*

The laboratory director noted that recommendations based on clinical outcome studies do not exist. The laboratory director consulted clinicians with expertise in the interpretation of coagulation tests, and their consensus judgment was that a CD of 15% was the standard of care.

- *Number of Runs, Replicates, and Range Rejection Limit*

To estimate the CD, one multiplies the acceptability criterion by the initial value from Analyzer A: $CD = 14.2 \times 0.15 = 2.13$. The following ratios are then calculated: CD/ Pooled $ST = 2.13/0.33 = 6.45$ and $SR/ST = 0.32/0.33 = 0.97$. Consulting Table B1, a value for $CD/ST \geq 4.00$ requires one run and one replicate. The range rejection limit is calculated as $0.70 \times CD = 0.70 \times 2.13 = 1.49$.

- *Comparison Data*

The samples were analyzed with the following results:

Replicate	Analyzer A	Analyzer B
1	14.2	15.1

Statistic	Sample 1
Range	$15.10 - 14.20 =$ **0.90**
Expected Difference	**1.42**
Absolute Adjusted Range	$\lvert 0.9 - 1.42 \rvert =$ **0.52***
Range Rejection Limit	**1.49**
Status	**Pass**

* Note that the range was adjusted by subtracting the expected difference and taking the absolute value of the results.

- *Conclusion*

Because the absolute value of the adjusted range was less than the range rejection limit, the comparability of the methods was considered acceptable, given the known underlying bias between the analyzers.

Appendix B. Tables of Runs, Replicates, and Range Rejection Limits

The following tables should be used to determine the number of runs, the number of replicates, and the range rejection limits to be applied to a comparison evaluation. The range rejection limits in the table have been calculated assuming an alpha error (Type I error) of 0.05 and a power of at least 80% (ie, Type II error of ≤ 0.2). The following entry in the "Number of Replicates" column: "—", indicates that the comparison cannot be conducted with the acceptability criterion that was selected. Select a less stringent acceptability criterion, recalculate the critical difference (CD) and CD/ST value and consult the table again. The same process must be followed if a within-run SD/total SD (SR/ST) value is not present in the table for a given CD/ST ratio.

Start with the table that corresponds to the number of measurement systems that will be compared (ie, two through 10). Find the column that represents the number of runs ("Runs") and first review the one run scenario. In the column of the ratio of CD to total SD (CD/ST) find the rows that best correspond to this ratio from the analyte to be tested. In the column of the ratio of within-run SD to total SD (SR/ST) find the single row that best corresponds to this ratio from the analyte to be tested. From this row, read the "Number of Replicates per System" column. If the table does not have a row corresponding to the above criteria or if the number of replicates is not feasible, then perform the same table look-up steps for the two run scenario. Once the best row is determined, perform the calculation specified in the "Range Rejection Limit" column to compute this limit from the CD.

The tables do not cover all possible values of the CD/ST or SR/ST ratios, so sometimes it will be necessary to interpolate table entries. Straightforward linear interpolation of the number of replicates should not be used, but linear interpolation of the inverse of the number of replicates is effective.

If the SR/ST value needed is close to one of the table entries, but the CD/ST ratio is not, then interpolation between CD/ST values is needed. Find the nearest table line whose SR/ST ratio matches the value needed and whose CD/ST ratio is smaller than the value needed. Write "x_0" for the CD/ST ratio of that line and "n_0" for the number of replicates listed on that line. Find the nearest table line whose SR/ST value matches the value needed and whose CD/ST ratio is larger than the value needed. Write "x_1" for the CD/ST ratio of that line, and "n_1" for the number of replicates. Write "x" for the CD/ST value whose number of replicates is needed. Then, the number of replicates needed is given by the following equation:

$$\frac{1}{n} = \frac{1}{n_0} + (x - x_0)\frac{n_0 - n_1}{n_0 n_1 (x_1 - x_0)}$$

For example, if one plans to use two runs to test five methods and have CD/ST = 2.2, SR/ST = 0.70, there are table entries for SR/ST = 0.70, but not for CD/ST = 2.2. The nearest table entries are:

$x_0 = 2.00$, with $n_0 = 17$
$x_1 = 2.50$, with $n_1 = 2$

To get the required number of replicates *n,* we need to solve:

$$\frac{1}{n} = \frac{1}{17} + (2.20 - 2.00)\frac{17 - 2}{17 * 2(2.50 - 2.00)} = 0.235,$$

giving $n = 1/0.235 = 4.25$.

Appendix B. (Continued)

	CD/ST		SR/ST		Replicates	
Table Location	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**
Lower Tabulated	x_0	2.0	NA	0.70	n_0	17
Upper Tabulated	x_1	2.5	NA	0.70	n_1	2
Interpolated	x	2.2	NA	0.70	n	**4.25**

Abbreviations: CD, critical difference; SR, within-run standard deviation; ST, total standard deviation.

To be conservative, round this up to 5.

Once the number of replicates is fixed, one will need to determine the appropriate range rejection limit. This can be found using the same inverse interpolation approach. Writing "L_0" and "L_1" for the range rejection limits next to the n_0 and n_1 entries. Then, use the range rejection limit "L" defined by:

$$\frac{1}{L} = \frac{1}{L_0} + (x - x_0)\frac{L_0 - L_1}{L_0 L_1 (x_1 - x_0)}.$$

Continuing with the example, the table entries for L_0 and L_1 are:

$x_0 = 2.00$, with $L_0 = 1.01$
$x_1 = 2.50$, with $L_1 = 0.95$

To get the required range rejection limit, one needs to solve:

$$\frac{1}{L} = \frac{1}{1.01} + (2.20 - 2.00)\frac{1.01 - 0.95}{1.01 * 0.95(2.50 - 2.00)} = 1.015,$$

giving $L = 1/1.015 = 0.99$.

	CD/ST		SR/ST		Replicates		Rejection Limit	
Table Location	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**
Lower Tabulated	x_0	2.0	NA	0.70	n_0	17	L_0	1.01
Upper Tabulated	x_1	2.5	NA	0.70	n_1	2	L_1	0.95
Interpolated	x	2.2	NA	0.70	n	4.25	L	**0.99**

Abbreviations: CD, critical difference; SR, within-run standard deviation; ST, total standard deviation.

The same approach can be used if the table has entries for the CD/ST needed, but not the SR/ST. In the general equation, just use SR/ST in place of CD/ST.

For example, using two runs to test five methods, where one has CD/ST = 2.00 and SR/ST = 0.73, use the following table entries:

SR/ST = x_0 = 0.70, n_0 = 17; SR/ST = x_1 = 0.75, n_1 = 6, and solve:

$$\frac{1}{n} = \frac{1}{17} + (0.73 - 0.70)\frac{17 - 6}{17 \bullet 6(0.75 - 0.70)} = 0.124,$$

giving $n = 8.1$.

Appendix B. (Continued)

	CD/ST		SR/ST		Replicates	
Table Location	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**
Lower Tabulated	NA	2.0	x_0	0.7	n_0	17
Upper Tabulated	NA	2.0	x_1	0.75	n_1	6
Interpolated	NA	2.0	x	0.73	n	**8.10**

Abbreviations: CD, critical difference; SR, within-run standard deviation; ST, total standard deviation.

To get the corresponding range rejection limit, use the following table entries:

SR/ST = x0 = 0.70, L0 = 1.01; SR/ST = x1 = 0.75, L1 = 1.00, and solve:

$$\frac{1}{L} = \frac{1}{1.01} + (0.73 - 0.70)\frac{1.01 - 1.00}{1.01 \bullet 1.00(0.75 - 0.70)} = 1.00,$$

giving L = 1.00 CD.

	CD/ST		SR/ST		Replicates		Rejection Limit	
Table Location	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**
Lower Tabulated	NA	2.0	x_0	0.7	n_0	17	L_0	1.01
Upper Tabulated	NA	2.0	x_1	0.75	n_1	6	L_1	1.00
Interpolated	NA	2.0	x	0.73	n	8.10	L	**1.00**

Abbreviations: CD, critical difference; SR, within-run standard deviation; ST, total standard deviation.

In this case, because the values of "L_0" and "L_1" were effectively identical, there is little need for a formal calculation.

Finally, if the table does not have the entries for either CD/ST or SR/ST, then one will need three stepwise interpolations. This can best be explained by example. To get the required number of replicates with two runs for CD/ST = 2.20, SR/ST = 0.73, perform the following three steps:

1) Use interpolation to get the required n for CD/ST = 2.2, SR/ST = 0.75, but do not round to a whole number. This gives n = 3.33 and a range rejection limit of 0.97.

	CD/ST		SR/ST		Replicates		Rejection Limit	
Table Location	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**
Lower Tabulated	x_0	2.0	NA	0.75	n_0	6	L_0	1.00
Upper Tabulated	x_1	2.5	NA	0.75	n_1	2	L_1	0.93
Interpolated	x	2.2	NA	0.75	N	**3.33**	L	**0.97**

Abbreviations: CD, critical difference; SR, within-run standard deviation; ST, total standard deviation.

2) Use interpolation to get the required n for CD/ST = 2.2, SR/ST = 0.70, again without rounding to a whole number. This gives n = 4.25 and a range rejection limit of 0.99.

Appendix B. (Continued)

	CD/ST		SR/ST		Replicates		Rejection Limit	
Table Location	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**
Lower Tabulated	x_0	2.0	NA	0.70	n_0	17	L_0	1.01
Upper Tabulated	x_1	2.5	NA	0.70	n_1	2	L_1	0.95
Interpolated	x	2.2	NA	0.70	n	**4.25**	L	**0.99**

Abbreviations: CD, critical difference; SR, within-run standard deviation; ST, total standard deviation.

3) Use interpolation of these two calculated values for CD/ST = 2.2, SR/ST = 0.70 and SR/ST = 0.75 to get the interpolated value for SR/ST = 0.73. This gives n = 3.65, which you would round upward to 4 and a range rejection limit of 0.98.

	CD/ST		SR/ST		Replicates		Rejection Limit	
Table Location	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**	**Variable**	**Value**
Lower Calculated	NA	2.2	x_0	0.70	n_0	4.25	L_0	0.99
Upper Calculated	NA	2.2	x_1	0.75	n_1	3.33	L_1	0.97
Interpolated	NA	2.2	x	0.73	n	**3.65**	L	**0.98**

Abbreviations: CD, critical difference; SR, within-run standard deviation; ST, total standard deviation.

Tables of Range Rejection Limits

Table B1. Number of Replicates per Run for a Two Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	16	0.70•CD	0.802	0.802
1	1.00	0.95	—	—	—	—
1	1.25	1.00	11	0.67•CD	0.833	0.833
1	1.25	0.95	444	0.70•CD	0.800	0.800
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	0.70•CD	0.800	0.800
1	1.50	0.95	20	0.70•CD	0.800	0.800
1	1.50	0.90	—	—	—	—
1	1.75	1.00	6	0.65•CD	0.856	0.856
1	1.75	0.95	10	0.69•CD	0.812	0.812
1	1.75	0.90	160	0.70•CD	0.800	0.800
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	0.70•CD	0.802	0.802
1	2.00	0.95	6	0.70•CD	0.803	0.803
1	2.00	0.90	13	0.70•CD	0.801	0.801
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	0.65•CD	0.858	0.858
1	2.50	0.95	4	0.64•CD	0.869	0.869
1	2.50	0.90	4	0.70•CD	0.801	0.801
1	2.50	0.85	6	0.70•CD	0.800	0.800

Appendix B. (Continued)

Table B1. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	2.50	0.80	17	0.70•CD	0.800	0.800
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	0.66•CD	0.846	0.846
1	3.00	0.95	2	0.69•CD	0.813	0.813
1	3.00	0.90	3	0.63•CD	0.876	0.876
1	3.00	0.85	3	0.67•CD	0.835	0.835
1	3.00	0.80	4	0.67•CD	0.834	0.834
1	3.00	0.75	5	0.69•CD	0.812	0.812
1	3.00	0.70	8	0.70•CD	0.800	0.800
1	3.00	0.60	—	—	—	—
1	4.00	—-	1	0.70•CD	0.802	0.802
2	1.00	1.00	8	0.70•CD	0.802	0.802
2	1.00	0.95	31	0.70•CD	0.800	0.800
2	1.00	0.90	—	—	—	—
2	1.25	1.00	6	0.65•CD	0.858	0.858
2	1.25	0.95	9	0.70•CD	0.800	0.800
2	1.25	0.90	90	0.70•CD	0.800	0.800
2	1.25	0.85	—	—	—	—
2	1.50	1.00	4	0.66•CD	0.846	0.846
2	1.50	0.95	5	0.69•CD	0.811	0.811
2	1.50	0.90	9	0.70•CD	0.802	0.802
2	1.50	0.85	79	0.70•CD	0.800	0.800
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	0.65•CD	0.856	0.856
2	1.75	0.95	4	0.64•CD	0.866	0.866
2	1.75	0.90	5	0.67•CD	0.835	0.835
2	1.75	0.85	7	0.70•CD	0.803	0.803
2	1.75	0.80	22	0.70•CD	0.800	0.800
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	0.70•CD	0.802	0.802
2	2.00	0.95	3	0.62•CD	0.886	0.886
2	2.00	0.90	3	0.67•CD	0.835	0.835
2	2.00	0.85	4	0.67•CD	0.835	0.835
2	2.00	0.80	5	0.69•CD	0.813	0.813
2	2.00	0.75	8	0.70•CD	0.800	0.800
2	2.00	0.70	—	—	—	—
2	2.50	1.00	2	0.56•CD	0.940	0.940
2	2.50	0.95	2	0.59•CD	0.917	0.917
2	2.50	0.90	2	0.61•CD	0.897	0.897
2	2.50	0.85	2	0.63•CD	0.876	0.876
2	2.50	0.80	2	0.65•CD	0.856	0.856
2	2.50	0.75	2	0.67•CD	0.835	0.835
2	2.50	0.70	2	0.69•CD	0.814	0.814
2	2.50	0.60	3	0.69•CD	0.813	0.813
2	2.50	0.50	6	0.70•CD	0.800	0.800
2	2.50	0.40	—	—	—	—

Appendix B. (Continued)

Table B1. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	3.00	—	1	0.66•CD	0.846	0.846
2	4.00	—	1	0.49•CD	0.979	0.979

Abbreviations: CD, constant difference; SR, within-run standard deviation; ST, total standard deviation.

Table B2. Number of Replicates per Run for a Three Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	15	0.86•CD	0.803	0.682
1	1.00	0.95	—	—	—	—
1	1.25	1.00	10	0.84•CD	0.822	0.704
1	1.25	0.95	108	0.87•CD	0.794	0.671
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	0.84•CD	0.823	0.704
1	1.50	0.95	17	0.86•CD	0.803	0.682
1	1.50	0.90	—	—	—	—
1	1.75	1.00	5	0.85•CD	0.813	0.693
1	1.75	0.95	9	0.85•CD	0.814	0.693
1	1.75	0.90	46	0.87•CD	0.794	0.671
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	0.83•CD	0.832	0.715
1	2.00	0.95	6	0.83•CD	0.832	0.715
1	2.00	0.90	10	0.87•CD	0.794	0.671
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	0.77•CD	0.886	0.783
1	2.50	0.95	3	0.84•CD	0.823	0.704
1	2.50	0.90	4	0.84•CD	0.823	0.705
1	2.50	0.85	5	0.87•CD	0.794	0.671
1	2.50	0.80	11	0.86•CD	0.803	0.682
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	0.79•CD	0.870	0.761
1	3.00	0.95	2	0.82•CD	0.841	0.726
1	3.00	0.90	2	0.86•CD	0.804	0.682
1	3.00	0.85	3	0.80•CD	0.860	0.749
1	3.00	0.80	3	0.84•CD	0.823	0.704
1	3.00	0.75	4	0.85•CD	0.814	0.693
1	3.00	0.70	5	0.87•CD	0.794	0.671
1	3.00	0.60	—	—	—	—
1	4.00	—	1	0.83•CD	0.832	0.715
2	1.00	1.00	8	0.83•CD	0.832	0.715
2	1.00	0.95	24	0.87•CD	0.794	0.671
2	1.00	0.90	—	—	—	—
2	1.25	1.00	5	0.84•CD	0.822	0.704
2	1.25	0.95	8	0.86•CD	0.803	0.682
2	1.25	0.90	38	0.87•CD	0.794	0.671
2	1.25	0.85	—	—	—	—

Appendix B. (Continued)

Table B2. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	1.50	1.00	4	0.79•CD	0.870	0.761
2	1.50	0.95	5	0.83•CD	0.833	0.715
2	1.50	0.90	8	0.85•CD	0.813	0.693
2	1.50	0.85	27	0.87•CD	0.794	0.671
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	0.78•CD	0.878	0.772
2	1.75	0.95	3	0.85•CD	0.813	0.693
2	1.75	0.90	4	0.84•CD	0.822	0.704
2	1.75	0.85	6	0.85•CD	0.813	0.693
2	1.75	0.80	12	0.87•CD	0.794	0.671
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	0.83•CD	0.832	0.715
2	2.00	0.95	3	0.74•CD	0.910	0.816
2	2.00	0.90	3	0.80•CD	0.860	0.749
2	2.00	0.85	3	0.85•CD	0.813	0.693
2	2.00	0.80	4	0.85•CD	0.813	0.693
2	2.00	0.75	6	0.86•CD	0.804	0.682
2	2.00	0.70	16	0.87•CD	0.794	0.671
2	2.00	0.60	—	—	—	—
2	2.50	1.00	2	0.67•CD	0.957	0.891
2	2.50	0.95	2	0.70•CD	0.939	0.861
2	2.50	0.90	2	0.73•CD	0.919	0.829
2	2.50	0.85	2	0.75•CD	0.902	0.805
2	2.50	0.80	2	0.78•CD	0.878	0.772
2	2.50	0.75	2	0.80•CD	0.860	0.749
2	2.50	0.70	2	0.82•CD	0.842	0.727
2	2.50	0.60	2	0.85•CD	0.813	0.693
2	2.50	0.50	3	0.86•CD	0.803	0.682
2	2.50	0.40	22	0.87•CD	0.794	0.671
2	2.50	0.30	—	—	—	—
2	3.00	—	1	0.79•CD	0.870	0.761
2	4.00	—	1	0.59•CD	0.988	0.955

Abbreviations: CD, constant difference; SR, within-run standard deviation; ST, total standard deviation.

Appendix B. (Continued)

Table B3. Number of Replicates per Run for a Four Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	15	0.94•CD	0.807	0.621
1	1.00	0.95	—	—	—	—
1	1.25	1.00	10	0.92•CD	0.825	0.642
1	1.25	0.95	104	0.95•CD	0.798	0.610
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	0.92•CD	0.825	0.641
1	1.50	0.95	17	0.94•CD	0.807	0.621
1	1.50	0.90	—	—	—	—
1	1.75	1.00	5	0.93•CD	0.816	0.631
1	1.75	0.95	9	0.93•CD	0.816	0.631
1	1.75	0.90	45	0.95•CD	0.798	0.610
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	0.91•CD	0.834	0.653
1	2.00	0.95	6	0.91•CD	0.834	0.652
1	2.00	0.90	10	0.95•CD	0.798	0.610
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	0.84•CD	0.892	0.731
1	2.50	0.95	3	0.92•CD	0.825	0.642
1	2.50	0.90	4	0.92•CD	0.825	0.641
1	2.50	0.85	5	0.95•CD	0.798	0.610
1	2.50	0.80	10	0.95•CD	0.798	0.610
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	0.86•CD	0.877	0.708
1	3.00	0.95	2	0.90•CD	0.843	0.663
1	3.00	0.90	2	0.94•CD	0.807	0.620
1	3.00	0.85	3	0.88•CD	0.860	0.685
1	3.00	0.80	3	0.92•CD	0.825	0.642
1	3.00	0.75	4	0.93•CD	0.816	0.630
1	3.00	0.70	5	0.95•CD	0.798	0.610
1	3.00	0.60	—	—	—	—
1	4.00	—	1	0.91•CD	0.834	0.653
2	1.00	1.00	8	0.91•CD	0.834	0.653
2	1.00	0.95	24	0.95•CD	0.798	0.610
2	1.00	0.90	—	—	—	—
2	1.25	1.00	5	0.92•CD	0.825	0.642
2	1.25	0.95	8	0.95•CD	0.798	0.609
2	1.25	0.90	37	0.95•CD	0.798	0.610
2	1.25	0.85	—	—	—	—
2	1.50	1.00	4	0.86•CD	0.877	0.708
2	1.50	0.95	5	0.91•CD	0.834	0.652
2	1.50	0.90	7	0.95•CD	0.798	0.610
2	1.50	0.85	26	0.95•CD	0.798	0.610
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	0.85•CD	0.885	0.720
2	1.75	0.95	3	0.93•CD	0.816	0.631
2	1.75	0.90	4	0.92•CD	0.825	0.642
2	1.75	0.85	6	0.93• CD	0.816	0.631

Appendix B. (Continued)

Table B3. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	1.75	0.80	12	0.95•CD	0.798	0.610
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	0.91•CD	0.834	0.653
2	2.00	0.95	3	0.82•CD	0.909	0.755
2	2.00	0.90	3	0.88•CD	0.861	0.685
2	2.00	0.85	3	0.93•CD	0.816	0.631
2	2.00	0.80	4	0.93•CD	0.816	0.631
2	2.00	0.75	6	0.94•CD	0.807	0.620
2	2.00	0.70	15	0.95•CD	0.798	0.610
2	2.00	0.60	—	—	—	—
2	2.50	1.00	2	0.73•CD	0.962	0.856
2	2.50	0.95	2	0.77•CD	0.942	0.812
2	2.50	0.90	2	0.80•CD	0.922	0.778
2	2.50	0.85	2	0.83•CD	0.901	0.743
2	2.50	0.80	2	0.85•CD	0.885	0.720
2	2.50	0.75	2	0.88•CD	0.861	0.685
2	2.50	0.70	2	0.90 •CD	0.843	0.663
2	2.50	0.60	2	0.94•CD	0.807	0.620
2	2.50	0.50	3	0.94•CD	0.807	0.621
2	2.50	0.40	17	0.95•CD	0.798	0.610
2	2.50	0.30	—	—	—	—
2	3.00	—	1	0.86•CD	0.877	0.708
2	4.00	—	1	0.65•CD	0.989	0.932

Abbreviations: CD, constant difference; SR, within-run standard deviation; ST, total standard deviation.

Appendix B. (Continued)

Table B4. Number of Replicates per Run for a Five Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	15	1.00•CD	0.801	0.571
1	1.00	0.95	—	—	—	—
1	1.25	1.00	10	0.98•CD	0.819	0.591
1	1.25	0.95	117	1.01•CD	0.792	0.560
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	0.98•CD	0.819	0.590
1	1.50	0.95	17	1.00•CD	0.801	0.571
1	1.50	0.90	—	—	—	—
1	1.75	1.00	5	0.99•CD	0.810	0.581
1	1.75	0.95	9	0.99•CD	0.810	0.580
1	1.75	0.90	50	1.01•CD	0.792	0.560
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	0.97•CD	0.828	0.601
1	2.00	0.95	6	0.97•CD	0.828	0.600
1	2.00	0.90	11	1.00•CD	0.801	0.569
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	0.90•CD	0.886	0.678
1	2.50	0.95	3	0.98•CD	0.819	0.591
1	2.50	0.90	4	0.97•CD	0.828	0.602
1	2.50	0.85	6	0.98•CD	0.819	0.590
1	2.50	0.80	11	1.00•CD	0.801	0.571
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	0.91•CD	0.877	0.667
1	3.00	0.95	2	0.96•CD	0.836	0.611
1	3.00	0.90	2	1.00•CD	0.801	0.570
1	3.00	0.85	3	0.93•CD	0.862	0.644
1	3.00	0.80	3	0.98•CD	0.819	0.591
1	3.00	0.75	4	0.98•CD	0.819	0.591
1	3.00	0.70	6	0.99•CD	0.810	0.582
1	3.00	0.60	—	—	—	—
1	4.00	—	1	0.97•CD	0.828	0.601
2	1.00	1.00	8	0.97•CD	0.828	0.601
2	1.00	0.95	25	1.00•CD	0.801	0.571
2	1.00	0.90	—	—	—	—
2	1.25	1.00	5	0.98•CD	0.819	0.591
2	1.25	0.95	8	1.01•CD	0.792	0.560
2	1.25	0.90	40	1.01•CD	0.792	0.560
2	1.25	0.85	—	—	—	—
2	1.50	1.00	4	0.91•CD	0.877	0.667
2	1.50	0.95	5	0.96•CD	0.836	0.612
2	1.50	0.90	8	0.99•CD	0.810	0.580
2	1.50	0.85	29	1.01•CD	0.792	0.559
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	0.90•CD	0.885	0.678
2	1.75	0.95	3	0.99•CD	0.810	0.580
2	1.75	0.90	4	0.98•CD	0.819	0.591
2	1.75	0.85	6	0.99•CD	0.810	0.580

Appendix B. (Continued)

Table B4. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	1.75	0.80	13	1.00 • CD	0.801	0.571
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	0.97 • CD	0.828	0.601
2	2.00	0.95	3	0.87 • CD	0.908	0.712
2	2.00	0.90	3	0.93 • CD	0.862	0.644
2	2.00	0.85	3	0.99 • CD	0.810	0.580
2	2.00	0.80	4	0.99 • CD	0.810	0.580
2	2.00	0.75	6	1.00 • CD	0.801	0.570
2	2.00	0.70	17	1.01 • CD	0.792	0.560
2	2.00	0.60	—	—	—	—
2	2.50	1.00	2	0.78 • CD	0.961	0.818
2	2.50	0.95	2	0.81 • CD	0.945	0.783
2	2.50	0.90	2	0.85 • CD	0.922	0.736
2	2.50	0.85	2	0.88 • CD	0.901	0.701
2	2.50	0.80	2	0.90 • CD	0.885	0.678
2	2.50	0.75	2	0.93 • CD	0.862	0.644
2	2.50	0.70	2	0.95 • CD	0.845	0.623
2	2.50	0.60	2	0.99 • CD	0.810	0.581
2	2.50	0.50	3	1.00 • CD	0.801	0.571
2	2.50	0.40	74	1.01 • CD	0.792	0.560
2	2.50	0.30	—	—	—	—
2	3.00	—	1	0.91 • CD	0.877	0.667
2	4.00	—	1	0.69 • CD	0.990	0.912

Abbreviations: CD, constant difference; SR, within-run standard deviation; ST, total standard deviation.

Appendix B. (Continued)

Table B5. Number of Replicates per Run for a Six Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	16	1.01•CD	0.826	0.568
1	1.00	0.95	—	—	—	—
1	1.25	1.00	10	1.02•CD	0.817	0.559
1	1.25	0.95	138	1.04•CD	0.800	0.539
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	1.02•CD	0.817	0.557
1	1.50	0.95	18	1.04•CD	0.800	0.537
1	1.50	0.90	—	—	—	—
1	1.75	1.00	5	1.03•CD	0.809	0.549
1	1.75	0.95	9	1.03•CD	0.808	0.547
1	1.75	0.90	59	1.04•CD	0.800	0.539
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	1.01•CD	0.826	0.568
1	2.00	0.95	6	1.01•CD	0.826	0.567
1	2.00	0.90	11	1.04•CD	0.800	0.538
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	0.94•CD	0.883	0.641
1	2.50	0.95	3	1.02•CD	0.817	0.558
1	2.50	0.90	4	1.01•CD	0.826	0.569
1	2.50	0.85	6	1.02•CD	0.817	0.558
1	2.50	0.80	12	1.04•CD	0.800	0.538
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	0.95•CD	0.875	0.632
1	3.00	0.95	2	1.00•CD	0.834	0.578
1	3.00	0.90	2	1.04•CD	0.800	0.538
1	3.00	0.85	3	0.97•CD	0.859	0.610
1	3.00	0.80	3	1.02•CD	0.817	0.558
1	3.00	0.75	4	1.03•CD	0.808	0.546
1	3.00	0.70	6	1.04•CD	0.800	0.537
1	3.00	0.60	—	—	—	—
1	4.00	—	1	1.01•CD	0.826	0.568
2	1.00	1.00	8	1.01•CD	0.826	0.568
2	1.00	0.95	26	1.04•CD	0.800	0.538
2	1.00	0.90	—	—	—	—
2	1.25	1.00	5	1.02•CD	0.817	0.559
2	1.25	0.95	9	1.02•CD	0.817	0.557
2	1.25	0.90	45	1.04•CD	0.800	0.539
2	1.25	0.85	—	—	—	—
2	1.50	1.00	4	0.95•CD	0.875	0.632
2	1.50	0.95	5	1.01•CD	0.826	0.566
2	1.50	0.90	8	1.03•CD	0.809	0.548
2	1.50	0.85	33	1.04•CD	0.800	0.539
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	0.95•CD	0.875	0.630
2	1.75	0.95	3	1.03•CD	0.809	0.548
2	1.75	0.90	4	1.03•CD	0.808	0.546
2	1.75	0.85	6	1.03•CD	0.809	0.548

Appendix B. (Continued)

Table B5. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	1.75	0.80	14	1.04•CD	0.800	0.539
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	1.01•CD	0.826	0.568
2	2.00	0.95	3	0.90•CD	0.911	0.689
2	2.00	0.90	3	0.97•CD	0.859	0.610
2	2.00	0.85	3	1.03•CD	0.809	0.548
2	2.00	0.80	4	1.03•CD	0.809	0.548
2	2.00	0.75	6	1.04•CD	0.800	0.539
2	2.00	0.70	22	1.04•CD	0.800	0.539
2	2.00	0.60	—	—	—	—
2	2.50	1.00	2	0.81•CD	0.962	0.795
2	2.50	0.95	2	0.85•CD	0.943	0.747
2	2.50	0.90	2	0.88•CD	0.925	0.712
2	2.50	0.85	2	0.92•CD	0.898	0.664
2	2.50	0.80	2	0.94•CD	0.883	0.643
2	2.50	0.75	2	0.97•CD	0.859	0.610
2	2.50	0.70	2	1.00•CD	0.834	0.576
2	2.50	0.60	2	1.04•CD	0.800	0.537
2	2.50	0.50	4	1.03•CD	0.809	0.548
2	2.50	0.40	—	—	—	—
2	3.00	—	1	0.95•CD	0.875	0.632
2	4.00	—	1	0.72•CD	0.990	0.894

Abbreviations: CD, critical difference; SR, within-run standard deviation; ST, total standard deviation.

Appendix B. (Continued)

Table B6. Number of Replicates per Run for a Seven Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	16	1.05•CD	0.816	0.531
1	1.00	0.95	—	—	—	—
1	1.25	1.00	10	1.06•CD	0.808	0.522
1	1.25	0.95	171	1.07•CD	0.800	0.515
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	1.06•CD	0.808	0.521
1	1.50	0.95	18	1.07•CD	0.799	0.514
1	1.50	0.90	—	—	—	—
1	1.75	1.00	5	1.07•CD	0.799	0.513
1	1.75	0.95	9	1.06•CD	0.808	0.524
1	1.75	0.90	71	1.07•CD	0.800	0.515
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	1.05•CD	0.816	0.531
1	2.00	0.95	6	1.04•CD	0.825	0.543
1	2.00	0.90	12	1.06•CD	0.808	0.524
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	0.97•CD	0.881	0.614
1	2.50	0.95	3	1.06•CD	0.808	0.521
1	2.50	0.90	4	1.05•CD	0.816	0.532
1	2.50	0.85	6	1.06•CD	0.808	0.521
1	2.50	0.80	13	1.07•CD	0.799	0.514
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	0.99•CD	0.866	0.593
1	3.00	0.95	2	1.03•CD	0.834	0.553
1	3.00	0.90	3	0.95•CD	0.896	0.637
1	3.00	0.85	3	1.01•CD	0.850	0.571
1	3.00	0.80	3	1.06•CD	0.808	0.521
1	3.00	0.75	4	1.06•CD	0.808	0.523
1	3.00	0.70	6	1.07•CD	0.800	0.515
1	3.00	0.60	—	—	—	—
1	4.00	—	1	1.05•CD	0.816	0.531
2	1.00	1.00	8	1.05•CD	0.816	0.531
2	1.00	0.95	27	1.07•CD	0.799	0.514
2	1.00	0.90	—	—	—	—
2	1.25	1.00	5	1.06•CD	0.808	0.522
2	1.25	0.95	9	1.05•CD	0.817	0.533
2	1.25	0.90	52	1.07•CD	0.800	0.515
2	1.25	0.85	—	—	—	—
2	1.50	1.00	4	0.99•CD	0.866	0.593
2	1.50	0.95	5	1.04•CD	0.825	0.542
2	1.50	0.90	8	1.07•CD	0.799	0.511
2	1.50	0.85	39	1.07•CD	0.800	0.515
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	0.98•CD	0.874	0.603
2	1.75	0.95	3	1.07•CD	0.799	0.512
2	1.75	0.90	4	1.06•CD	0.808	0.523
2	1.75	0.85	6	1.07•CD	0.799	0.512

Appendix B. (Continued)

Table B6. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	1.75	0.80	15	1.07•CD	0.800	0.515
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	1.05•CD	0.816	0.531
2	2.00	0.95	3	0.94•CD	0.903	0.648
2	2.00	0.90	3	1.00•CD	0.858	0.584
2	2.00	0.85	3	1.07•CD	0.799	0.512
2	2.00	0.80	4	1.07•CD	0.799	0.512
2	2.00	0.75	7	1.07•CD	0.799	0.511
2	2.00	0.70	30	1.07•CD	0.800	0.515
2	2.00	0.60	—	—	—	—
2	2.50	1.00	2	0.84•CD	0.961	0.768
2	2.50	0.95	2	0.88•CD	0.941	0.720
2	2.50	0.90	2	0.91•CD	0.923	0.685
2	2.50	0.85	2	0.95•CD	0.896	0.637
2	2.50	0.80	2	0.98•CD	0.874	0.603
2	2.50	0.75	2	1.00•CD	0.858	0.584
2	2.50	0.70	2	1.03•CD	0.833	0.552
2	2.50	0.60	2	1.07•CD	0.799	0.514
2	2.50	0.50	4	1.07•CD	0.799	0.512
2	2.50	0.40	—	—	—	—
2	3.00	—	1	0.99•CD	0.866	0.593
2	4.00	—	1	0.74•CD	0.990	0.882

Abbreviations: CD, constant difference; SR, within-run standard deviation; ST, total standard deviation.

Appendix B. (Continued)

Table B7. Number of Replicates per Run for an Eight Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	16	1.08•CD	0.810	0.504
1	1.00	0.95	—	—	—	—
1	1.25	1.00	10	1.09•CD	0.802	0.496
1	1.25	0.95	219	1.10•CD	0.793	0.487
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	1.09•CD	0.802	0.494
1	1.50	0.95	19	1.09•CD	0.803	0.498
1	1.50	0.90	—	—	—	—
1	1.75	1.00	6	1.00•CD	0.876	0.588
1	1.75	0.95	9	1.09•CD	0.803	0.498
1	1.75	0.90	89	1.10•CD	0.793	0.487
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	1.08•CD	0.810	0.504
1	2.00	0.95	6	1.07•CD	0.819	0.516
1	2.00	0.90	12	1.09•CD	0.802	0.498
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	0.99•CD	0.883	0.599
1	2.50	0.95	3	1.09•CD	0.802	0.495
1	2.50	0.90	4	1.08•CD	0.811	0.505
1	2.50	0.85	6	1.09•CD	0.802	0.495
1	2.50	0.80	14	1.10•CD	0.793	0.486
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	1.02•CD	0.860	0.563
1	3.00	0.95	2	1.06•CD	0.828	0.526
1	3.00	0.90	3	0.97•CD	0.898	0.621
1	3.00	0.85	3	1.03•CD	0.852	0.556
1	3.00	0.80	3	1.09•CD	0.802	0.495
1	3.00	0.75	4	1.09•CD	0.802	0.497
1	3.00	0.70	7	1.09•CD	0.803	0.498
1	3.00	0.60	—	—	—	—
1	4.00	—	1	1.08•CD	0.810	0.504
2	1.00	1.00	8	1.08•CD	0.810	0.504
2	1.00	0.95	28	1.10•CD	0.793	0.486
2	1.00	0.90	—	—	—	—
2	1.25	1.00	5	1.09•CD	0.802	0.496
2	1.25	0.95	9	1.08•CD	0.811	0.507
2	1.25	0.90	62	1.10•CD	0.793	0.487
2	1.25	0.85	—	—	—	—
2	1.50	1.00	4	1.02•CD	0.860	0.563
2	1.50	0.95	5	1.07•CD	0.819	0.515
2	1.50	0.90	8	1.10•CD	0.793	0.486
2	1.50	0.85	48	1.10•CD	0.793	0.487
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	1.00•CD	0.876	0.588
2	1.75	0.95	3	1.10•CD	0.793	0.487
2	1.75	0.90	4	1.09•CD	0.802	0.497
2	1.75	0.85	6	1.10•CD	0.793	0.486

Appendix B. (Continued)

Table B7. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	1.75	0.80	17	1.10•CD	0.793	0.486
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	1.08•CD	0.810	0.504
2	2.00	0.95	3	0.96•CD	0.905	0.631
2	2.00	0.90	3	1.03•CD	0.852	0.556
2	2.00	0.85	3	1.10•CD	0.793	0.486
2	2.00	0.80	4	1.10•CD	0.793	0.487
2	2.00	0.75	7	1.10•CD	0.793	0.486
2	2.00	0.70	48	1.10•CD	0.793	0.487
2	2.00	0.60	—	—	—	—
2	2.50	1.00	2	0.86•CD	0.961	0.751
2	2.50	0.95	2	0.90•CD	0.941	0.702
2	2.50	0.90	2	0.94•CD	0.918	0.654
2	2.50	0.85	2	0.97•CD	0.898	0.621
2	2.50	0.80	2	1.00•CD	0.876	0.588
2	2.50	0.75	2	1.03•CD	0.852	0.556
2	2.50	0.70	2	1.06•CD	0.828	0.524
2	2.50	0.60	3	1.06•CD	0.828	0.526
2	2.50	0.50	4	1.10•CD	0.793	0.487
2	2.50	0.40	—	—	—	—
2	3.00	—	1	1.02•CD	0.860	0.563
2	4.00	—	1	0.76•CD	0.990	0.868

Abbreviations: CD, constant difference; SR, within-run standard deviation; ST, total standard deviation.

Appendix B. (Continued)

Table B8. Number of Replicates per Run for a Nine Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	16	1.10•CD	0.810	0.489
1	1.00	0.95	—	—	—	—
1	1.25	1.00	10	1.11•CD	0.802	0.482
1	1.25	0.95	299	1.12•CD	0.793	0.470
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	1.11•CD	0.802	0.480
1	1.50	0.95	20	1.11•CD	0.802	0.479
1	1.50	0.90	—	—	—	—
1	1.75	1.00	6	1.03•CD	0.867	0.556
1	1.75	0.95	10	1.09•CD	0.819	0.498
1	1.75	0.90	116	1.12•CD	0.793	0.470
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	1.10•CD	0.810	0.489
1	2.00	0.95	6	1.10•CD	0.810	0.487
1	2.00	0.90	13	1.11•CD	0.801	0.478
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	1.02•CD	0.875	0.567
1	2.50	0.95	3	1.11•CD	0.802	0.481
1	2.50	0.90	4	1.10•CD	0.811	0.491
1	2.50	0.85	6	1.11•CD	0.802	0.481
1	2.50	0.80	16	1.11•CD	0.802	0.482
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	1.04•CD	0.859	0.546
1	3.00	0.95	2	1.09•CD	0.818	0.497
1	3.00	0.90	3	1.00•CD	0.890	0.588
1	3.00	0.85	3	1.06•CD	0.843	0.526
1	3.00	0.80	3	1.11•CD	0.802	0.481
1	3.00	0.75	4	1.12•CD	0.793	0.469
1	3.00	0.70	8	1.11•CD	0.802	0.480
1	3.00	0.60	—	—	—	—
1	4.00	—	1	1.10•CD	0.810	0.489
2	1.00	1.00	8	1.10•CD	0.810	0.489
2	1.00	0.95	29	1.12•CD	0.793	0.470
2	1.00	0.90	—	—	—	—
2	1.25	1.00	5	1.11•CD	0.802	0.482
2	1.25	0.95	9	1.11•CD	0.801	0.479
2	1.25	0.90	74	1.12•CD	0.793	0.470
2	1.25	0.85	—	—	—	—
2	1.50	1.00	4	1.04•CD	0.859	0.546
2	1.50	0.95	5	1.10•CD	0.810	0.486
2	1.50	0.90	9	1.10•CD	0.810	0.488
2	1.50	0.85	61	1.12•CD	0.793	0.470
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	1.03•CD	0.867	0.556
2	1.75	0.95	4	1.01•CD	0.882	0.579
2	1.75	0.90	4	1.12•CD	0.792	0.469
2	1.75	0.85	7	1.10•CD	0.810	0.488

Appendix B. (Continued)

Table B8. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	1.75	0.80	19	1.12•CD	0.793	0.469
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	1.10•CD	0.810	0.489
2	2.00	0.95	3	0.98•CD	0.904	0.613
2	2.00	0.90	3	1.06•CD	0.843	0.525
2	2.00	0.85	4	1.05•CD	0.852	0.539
2	2.00	0.80	5	1.09•CD	0.818	0.497
2	2.00	0.75	8	1.11•CD	0.802	0.480
2	2.00	0.70	105	1.12•CD	0.793	0.470
2	2.00	0.60	—	—	—	—
2	2.50	1.00	2	0.88•CD	0.960	0.732
2	2.50	0.95	2	0.92•CD	0.940	0.683
2	2.50	0.90	2	0.96•CD	0.917	0.635
2	2.50	0.85	2	1.00•CD	0.890	0.588
2	2.50	0.80	2	1.03•CD	0.867	0.556
2	2.50	0.75	2	1.06•CD	0.843	0.525
2	2.50	0.70	2	1.08•CD	0.827	0.509
2	2.50	0.60	3	1.09•CD	0.818	0.496
2	2.50	0.50	5	1.11•CD	0.802	0.482
2	2.50	0.40	—	—	—	—
2	3.00	—	1	1.04•CD	0.859	0.546
2	4.00	—	1	0.78•CD	0.990	0.852

Abbreviations: CD, constant difference; SR, within-run standard deviation; ST, total standard deviation.

Appendix B. (Continued)

Table B9. Number of Replicates per Run for a 10 Method Comparison

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
1	1.00	1.00	16	1.12•CD	0.807	0.473
1	1.00	0.95	—	—	—	—
1	1.25	1.00	11	1.08•CD	0.841	0.510
1	1.25	0.95	451	1.13•CD	0.799	0.465
1	1.25	0.90	—	—	—	—
1	1.50	1.00	7	1.13•CD	0.799	0.464
1	1.50	0.95	20	1.13•CD	0.799	0.463
1	1.50	0.90	—	—	—	—
1	1.75	1.00	6	1.05•CD	0.864	0.538
1	1.75	0.95	10	1.11•CD	0.816	0.482
1	1.75	0.90	162	1.13•CD	0.799	0.465
1	1.75	0.85	—	—	—	—
1	2.00	1.00	4	1.12•CD	0.807	0.473
1	2.00	0.95	6	1.12•CD	0.807	0.471
1	2.00	0.90	13	1.13•CD	0.798	0.462
1	2.00	0.85	—	—	—	—
1	2.50	1.00	3	1.04•CD	0.872	0.548
1	2.50	0.95	4	1.02•CD	0.887	0.571
1	2.50	0.90	4	1.13•CD	0.798	0.461
1	2.50	0.85	6	1.13•CD	0.799	0.465
1	2.50	0.80	17	1.13•CD	0.799	0.464
1	2.50	0.75	—	—	—	—
1	3.00	1.00	2	1.06•CD	0.856	0.528
1	3.00	0.95	2	1.11•CD	0.815	0.480
1	3.00	0.90	3	1.02•CD	0.887	0.568
1	3.00	0.85	3	1.08•CD	0.840	0.508
1	3.00	0.80	4	1.08•CD	0.840	0.509
1	3.00	0.75	5	1.11•CD	0.816	0.481
1	3.00	0.70	8	1.13•CD	0.799	0.464
1	3.00	0.60	—	—	—	—
1	4.00	—	1	1.12•CD	0.807	0.473
2	1.00	1.00	8	1.12•CD	0.807	0.473
2	1.00	0.95	31	1.13•CD	0.799	0.463
2	1.00	0.90	—	—	—	—
2	1.25	1.00	6	1.04•CD	0.872	0.548
2	1.25	0.95	9	1.13•CD	0.799	0.463
2	1.25	0.90	90	1.13•CD	0.799	0.465
2	1.25	0.85	—	—	—	—
2	1.50	1.00	4	1.06•CD	0.856	0.528
2	1.50	0.95	5	1.12•CD	0.807	0.470
2	1.50	0.90	9	1.12•CD	0.807	0.472
2	1.50	0.85	80	1.13•CD	0.799	0.465
2	1.50	0.80	—	—	—	—
2	1.75	1.00	3	1.05•CD	0.864	0.538
2	1.75	0.95	4	1.03•CD	0.879	0.560
2	1.75	0.90	5	1.08•CD	0.840	0.507
2	1.75	0.85	7	1.12•CD	0.807	0.472

Appendix B. (Continued)

Table B9. (Continued)

Runs	CD/ST	SR/ST	Number of Replicates	Range Rejection Limit	1 Method Out Power	Least Favorable Configuration Power
2	1.75	0.80	22	1.13 • CD	0.799	0.464
2	1.75	0.75	—	—	—	—
2	2.00	1.00	2	1.12 • CD	0.807	0.473
2	2.00	0.95	3	1.00 • CD	0.901	0.593
2	2.00	0.90	3	1.08 • CD	0.840	0.507
2	2.00	0.85	4	1.08 • CD	0.840	0.506
2	2.00	0.80	5	1.11 • CD	0.815	0.480
2	2.00	0.75	8	1.13 • CD	0.799	0.464
2	2.00	0.70	—	—	—	—
2	2.50	1.00	2	0.90 • CD	0.958	0.711
2	2.50	0.95	2	0.94 • CD	0.938	0.662
2	2.50	0.90	2	0.98 • CD	0.914	0.615
2	2.50	0.85	2	1.02 • CD	0.887	0.568
2	2.50	0.80	2	1.05 • CD	0.864	0.538
2	2.50	0.75	2	1.08 • CD	0.840	0.507
2	2.50	0.70	2	1.10 • CD	0.824	0.492
2	2.50	0.60	3	1.11 • CD	0.815	0.479
2	2.50	0.50	6	1.13 • CD	0.799	0.463
2	2.50	0.40	—	—	—	—
2	3.00	—	1	1.06 • CD	0.856	0.528
2	4.00	—	1	0.80 • CD	0.989	0.834

Abbreviations: CD, constant difference; SR, within-run standard deviation; ST, total standard deviation.

Appendix C. Statistical Concepts

C1 Hypothesis Testing

Hypothesis testing is a statistical tool for using data to make inferences or draw conclusions. For example, in this document, procedures are described for using hypothesis testing to draw conclusions about the comparability of two or more laboratory methods. For purposes of this discussion, *methods* are instruments or assay systems that may or may not be of the same make and model, or may or may not be in the same laboratory. In general terms, a hypothesis test has three components: 1) a *statistic* with a known, estimated, or assumed probability distribution; 2) a *hypothesis* about a population or situation represented by the statistic; and 3) a *critical value,* or decision limit, against which the statistic is compared to make an inference about the validity of the hypothesis.

The statistic in a hypothesis test is a number calculated from data obtained through observations, experiments, surveys, etc. For method comparison hypothesis tests, data are generated from analysis of specimens by two or more laboratory methods. The statistics of interest are related to ranges, means, and variances. Under sets of basic assumptions, the probability distributions of the statistics are well characterized.

A hypothesis is a simple statement about a situation of interest. For method comparisons, the hypotheses are statements such as *the means of the underlying populations of the datasets are equal*; *the difference between the population means is zero;* etc. The hypothesis to be tested is often referred to as the *null hypothesis.* Its converse is called the *alternative hypothesis.*

The *critical value* is a number (or pair of numbers) that defines the limit (or limits) beyond which it would be unlikely to obtain a value of the statistic if the null hypothesis is true. If the statistic is beyond (more extreme than) the critical value, the null hypothesis is rejected, or inferred not to be true. If the statistic does not exceed the critical value, the null hypothesis is not rejected, or inferred not to be false. The critical value is derived from two factors: 1) the probability distribution of the statistic; and 2) the significance level. The *significance level* is the probability of falsely (incorrectly) rejecting the null hypothesis when it is actually true. Commonly used, though somewhat arbitrary, significance levels are 5% and 1%.

Figure C1 illustrates the concepts of hypothesis testing. The upper curve illustrates the probability distribution of a statistic when the null hypothesis is true. The critical values are selected so the area of the shaded regions equals the significance level. If the value of the statistic generated by a method comparison study falls between the critical values, then the null hypothesis is not rejected. If the value of the statistic falls beyond the critical values, then the null hypothesis is rejected. The area of the shaded region equals the probability of falsely rejecting the null hypothesis. Note that in this figure, the probability distribution of the statistic is represented as normal. Many statistics have non-normal distributions, but the same concepts apply.

Appendix C. (Continued)

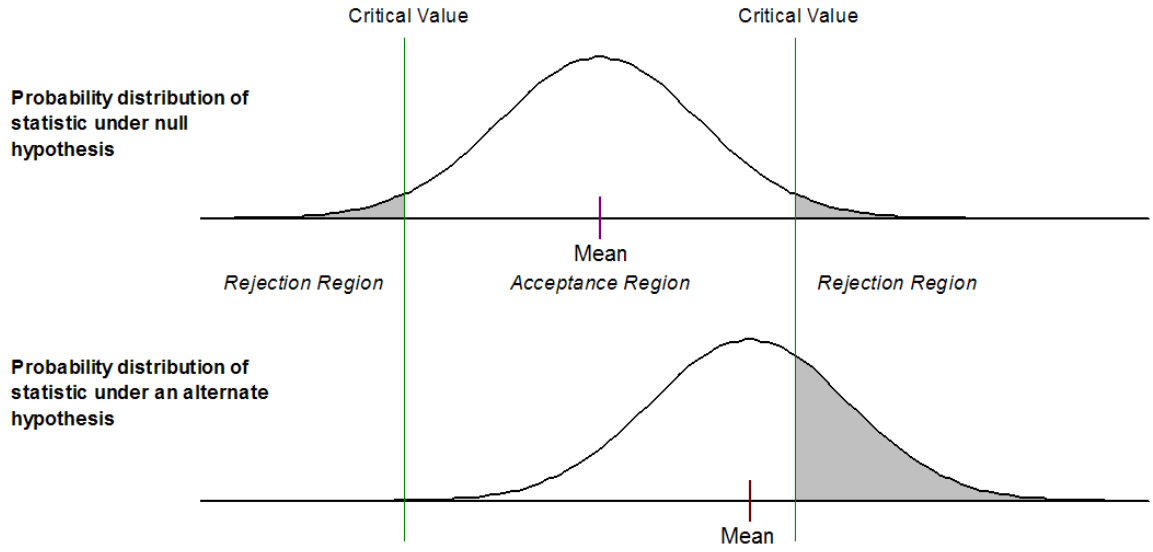

Figure C1. Illustration of Hypothesis Testing and Power Determination. See text for details.

C2 Power

Power is defined as the probability of correctly rejecting the null hypothesis when it is false. Power depends on both the critical value (which is related to the significance level, as discussed above) and the "degree of incorrectness" of the null hypothesis. Figure C1 illustrates both of these points. The upper curve shows the probability distribution of the statistic under the null hypothesis. The lower curve shows the probability distribution of the statistic under one instance of the alternative hypothesis. Note that the distribution of the statistic under the alternative hypothesis is "shifted," but the critical values do not change. Consequently, the probability of rejecting the null hypothesis is greater under the alternative hypothesis than under the null hypothesis. This probability is the *power*.

Power is affected by the critical values. If the significance level is changed so the critical values are moved farther into the tails of the distribution or closer to the center, the probability of rejection will decrease or increase, respectively, under both the null and alternative hypotheses. Thus, significance level and power change in the same direction.

Power is also affected by the "degree of incorrectness" of the null hypothesis. In the illustration, if the distribution of the statistic under the alternative hypothesis is shifted left, then power decreases; if it is shifted right, power increases. In other words, the greater the difference between the mean of the actual distribution (alternative hypothesis) and the assumed distribution (null hypothesis), the higher the probability of correctly rejecting the null hypothesis, or the greater the power.

It should be noted that power is also influenced by the variance of the test methods and the sample size. The probability of detecting a true difference between test methods is higher for more precise methods than for less precise methods, and power increases with the number of measurements used in the hypothesis test.

Appendix C. (Continued)

C3 Type I and Type II Errors

In hypothesis testing, two types of errors may be made. Type I error, or alpha error, occurs if the null hypothesis is rejected when it is actually true. Type II error, or beta error, occurs if the null hypothesis is accepted when it is actually false. The probability of a Type I error is controlled by the selection of the significance level, usually represented as "α." The probability of a Type II error is usually represented by "β" and is a function of four factors: the significance level (α), the degree of incorrectness of the null hypothesis, the variance of the test methods, and the sample size. These same factors influence power, as described above, and, actually, the power of a hypothesis test is $1-\beta$.

A hypothesis test may be designed to provide desired probabilities for Type I and Type II errors or, in other words, for desired significance levels and power. Usually, in hypothesis testing, the probability of Type I error is known because a particular significance level, α, is selected; but it is important to estimate β as well, to avoid running a hypothesis test with little probability of detecting important differences between methods.

Consider the following example: Suppose a specimen with an analyte concentration of 2.5 mmol/L is going to be tested by two methods, and because of the precision of the methods four replicates from each could detect a difference of 0.187 at a significance level of $\alpha = 0.05$. If the true difference is 0.187, the null hypothesis will be rejected only 50% of the time (power = $100(1-\beta)$). Now, if one wants the probability of β error to be only 0.10, or power of 90%, the number of replicates would need to be increased. Without going through all the calculations, by increasing replication the null hypothesis will be rejected if the test result's difference is greater than 0.139, and because the true difference is 0.187 the result's mean difference would have a 90% chance of being above 0.139.

C4 Range Test

The measurements most people think about when the field of statistics is mentioned are sample mean (or average) and SD. Such measurements are extremely powerful in characterizing a population, especially if the number of times this population is sampled is large. However, when the sample size is small, such measurements may not be the most efficient way to either characterize a population or to perform a hypothesis test.

A number of efficient statistical measurements called shortcut procedures[1] based on order statistics are available for instances in which only a small sample size is available. One such measurement is the range. In the case of comparing two or more measurement systems using multiple samples (replicates) there are nk observations X_{ij} ($i = 1,\ldots, n$; $j = 1,\ldots, k$) where n is the number of measurement systems and k is the number of replicates on each system. The range can be computed by first taking the mean $\overline{X}_i$ for each system over the k replicates. The range w is thus:

$$w = \overline{X}_{\max} - \overline{X}_{\min}$$

where $\overline{X}_{\max}$ is the maximum observed mean and $\overline{X}_{\min}$ is the minimum observed mean.

The null hypothesis for the range test is that the true measurement system means μ_i are all equal. The alternative hypothesis is that the range of the true measurement system means is greater than or equal to the critical difference (CD).

Appendix C. (Continued)

All potential configurations of n measurement systems can be described mathematically and the Type II error from a range test can thus be computed by integrating across all possible configurations of means of the measurement systems. However, a far more efficient, and practical, way to determine Type II error and thus sample size is to propose a configuration on which to perform these calculations.

The least favorable configuration[2] for a range equal to the CD (ie, that configuration resulting in the greatest potential Type II error) has a central set of measurement systems where all means are equal and two measurement systems with means equally far apart in opposite directions from this central set of means (μ − CD/2, μ, μ, …, μ, μ, μ + CD/2). The most favorable is where half of the system means are equal to μ − CD/2 and half are equal to μ + CD/2. A far more likely scenario (1 method out scenario) has a set of measurement systems where all but one system has the same mean and the remaining measurement system has a mean that is removed from this common mean; (μ − CD, μ, …, μ) or (μ, …, μ, μ + CD).

Once the configuration of the measurement system means ($\overline{X}_i$, noted below as X_i) is determined then given $X_i = \mu + \delta_i + Z_i$ with (i = 1,…, n) where:

$\sum_{i=1}^{n} \delta_i = 0$, and

Z_i has a normal distribution with a mean of zero and SD of 1, then W' = range X_i may be called the noncentral range. The cumulative distribution function (CDF) is given by the following:

$$\Pr(W' \le w) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(x_i - \delta_i) \left\{ \prod_{j=1, j \ne i}^{n} [\Phi(x_i - \delta_j + w) - \Phi(x_i - \delta_j)] \right\} dx_i$$

where $\phi(x)$ is the standard normal probability density function (PDF) and $\Phi(x)$ is the standard normal CDF. The power of each scenario can thus be calculated by determining these distributions given their underlying configurations. Through an iterative approach the range rejection limit (ie, the limit against which the measured range is compared to determine if the null hypothesis of equal means is rejected), plus the number of replicates required to achieve at least an 80% power to detect a true range of means equal to the CD can then be calculated.

"Shortcut procedures" such as the range are relatively easy to compute. The obvious difficulty, given the required computations above, is that power calculations cannot be provided through a simple calculation or a standard distribution. This is why the tables corresponding to a number of systems to be tested are provided in Appendix B. For each table, the replicate size is provided for the 1 method out scenario that gives at least an 80% power to detect a true range of measurement system means equal to the CD when the measured range is greater than the listed range rejection limit. In the last column of each table, the power to detect a range equal to the CD is also provided for the least favorable configuration scenario given the listed sample size.

C5 Within-run vs Total Standard Deviation

When imprecision is determined under typical study designs such as that proposed in CLSI document EP05,[3] the overall imprecision can be separated into within-run and between-run components. These are also called, respectively, repeatability and reproducibility. Within-run SD (SR or CV_w) is what is measured when multiple replicates are run within a short period of time. Between-run imprecision (S_b or CV_b) can be measured by first averaging over each such run and determining the variability across all the runs.

Appendix C. (Continued)

The reason the tables are structured to include both one run and two run options is that increasing replications in a single run does not decrease any inherent between-run variability. The only way to reduce such variability is to increase the number of runs, in this case to two runs. Such an option is necessary when a considerable proportion of the system imprecision is due to between-run imprecision. This corresponds to a low within-run to total SD ratio (SR/ST) as listed in the tables.

As a hypothetical example, assume a true instrument specific mean value of a sample is 100 and the within- and between-run imprecision is 5%. If you obtain a single result for a sample in one run, you will have a within-run bias and a between-run bias associated with that result, and on average the bias would be ±5% for each component of imprecision. If one were to test 100 replicates of that same sample in one run, one reduces the within-run uncertainty to $CV_w/\sqrt{N}$ = 5/10 = 0.5. One would then have a good estimate of the mean for that run. However, there is relatively high probability that the average of those 100 results would have a bias due to that run of ≈±5. The estimated total CV would be the following:

$$\sqrt{\frac{CV_w^2}{N}+CV_b^2}=\sqrt{\frac{25}{100}+25}=5.02.$$

Using the same number of tests, but testing 50 replicates of the same sample in run one and 50 in another run, the within-run uncertainty for each run would be $CV_w/\sqrt{2*50}$ and the between-run $CV_b/\sqrt{2}$. The estimated total CV would be the following:

$$\sqrt{\frac{CV_w^2}{(2*50)}+\frac{CV_b^2}{2}}=3.57.$$

Therefore, if between-run imprecision is a large enough component of the total imprecision, then splitting the same total number of tests into two runs will help to obtain an average result closer to the true instrument-specific mean for the sample.

References for Appendix C

1 David HA, Nagaraja HN. *Order Statistics.* 3rd ed. Hoboken, NJ: Wiley InterScience; 2003.

2 Chen SY, Chen HJ. A range test for the equivalency of means under unequal variances. *Technometrics.* 1999;41(3);250-260.

3 CLSI/NCCLS. *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition.* CLSI/NCCLS document EP05-A2. Wayne, PA: NCCLS; 2004.

Appendix D. Biological Variation

D1 Analytical Difference Between Two Results

Two results are said to be analytically different if the difference between them is more than could be accounted for by the combined analytical imprecision that may be present in both results. The total analytical imprecision (CV_{AT}) in two results using the same method having an imprecision of CV_A is defined as:

$$CV_{AT} = [CV_A^2 + CV_A^2]^{0.5}$$

$$CV_{AT} = [2CV_A^2]^{0.5}$$

$$CV_{AT} = [2]^{0.5} \bullet [CV_A^2]^{0.5}$$

$$CV_{AT} = 1.41 \bullet CV_A \quad (1)$$

In order to be 95% confident that such a combined imprecision has been exceeded, one needs to multiply CV_{AT} by a *z* value corresponding to a 95% probability (1.96) to derive the critical analytical difference (CD_A) that indicates a difference between two results that is greater than combined analytical imprecision.

$$CD_A = 1.96 \bullet 1.41 \bullet CV_A$$

$$CD_A = 2.77 \bullet CV_A \quad (2)$$

D2 Biological Difference Between Two Results

Two results are said to be biologically different if the difference between them is more than could be accounted for by both the combined analytical imprecision that may be present in both results as well as the combined biological variability that may occur in the parameter in a stable patient from day to day. The total analytical imprecision (CV_{AIT}) in two results using the same method having an imprecision of CV_A and a biological within-subject day-to-day variability of CV_I is defined as:

$$CV_{AIT} = [CV_A^2 + CV_I^2 + CV_A^2 + CV_I^2]^{0.5}$$

$$CV_{AIT} = [2CV_A^2 + 2CV_I^2]^{0.5} \quad (3)$$

$$CV_{AIT} = [2]^{0.5} \bullet [CV_A^2 + CV_I^2]^{0.5}$$

$$CV_{AIT} = 1.41 \bullet [CV_A^2 + CV_I^2]^{0.5}$$

In order to be 95% confident that such a combined imprecision has been exceeded, one needs to multiply this combined biological CV_{AIT} with a *z* value corresponding to a 95% probability (1.96) to derive the critical analytical difference (CD_{AI}) that indicates a difference between two results that is greater than combined analytical imprecision and within-subject biological variability.

$$CD_{AI} = 1.96 \bullet 1.41 \bullet [CV_A^2 + CV_I^2]^{0.5} \quad (4)$$

$$CD_{AI} = 2.77 \bullet [CV_A^2 + CV_I^2]^{0.5} \quad (5)$$

Appendix D. (Continued)

D3 Critical Biological Difference Between Two Results Being Performed With a Method of Desirable Analytical Imprecision

Desirable analytical imprecision has been defined by Harris[1] as well as Fraser and Petersen[2] as an imprecision that is less than half the within-subject biological variability.

$$CV_{Ad} \leq 0.5\ CV_I$$

If one assumes the method has desirable imprecision and substitutes this requirement into equation (4), the critical biologically significant change with desirable imprecision CV_{AdIT} becomes:

$$CV_{AdIT} = [CV_A^2 + CV_I^2 + CV_A^2 + CV_I^2]^{0.5}$$

$$CV_{AdIT} = [(0.5CV_I)^2 + CV_I^2 + (0.5CV_I)^2 + CV_I^2]^{0.5}$$

$$CV_{AdIT} = [0.25CV_I^2 + CV_I^2 + 0.25CV_I^2 + CV_I^2]^{0.5}$$

$$CV_{AdIT} = [2.5CV_I^2]^{0.5}$$

$$CV_{AdIT} = [2.5]^{0.05} \bullet [CV_I^2]^{0.5}$$

$$CV_{AdIT} = 1.58 \bullet CV_I$$

Similarly, in order to be 95% confident that such a combined imprecision has been exceeded, one needs to multiply this combined biological CV_{AdIT} with a *z* value corresponding to a 95% probability (1.96) to derive the critical analytical difference (CD_{AdI}) that indicates a difference between two results that is greater than combined, but desirable analytical imprecision.

$$CD_{AdI} = 1.96 \bullet 1.58 \bullet CV_I$$

$$CD_{AdI} = 3.10 \bullet CV_I \quad (6)$$

As analytical imprecision approaches, equation (5) becomes:

$$CD_I = 2.77 \bullet CV_I$$

A simple goal can be that the allowable bias between two methods ($|Bias_2 - Bias_1| = Bias_T$) for monitoring subjects is no more than the increase in the CD between two results attributable to the assay variability.

$$Bias_T \leq CD_{AI} - CD_I$$

Thus, a simple goal can be that the allowable bias between two *desirable* methods ($|Bias_{2d} - Bias_{1d}| = Bias_{Td}$) for monitoring subjects is:

$$Bias_{Td} \leq 3.10\,CV_I - 2.77\,CV_I = 0.33\,CV_I$$

References for Appendix D

[1] Harris EK. Statistical principles underlying analytical goal-setting in clinical chemistry. *Am J Clin Pathol.* 1979;72(2 Suppl):374-382.

[2] Fraser CG, Petersen PH. The importance of imprecision. *Ann Clin Biochem.* 1991;28(Pt 3):207-211.

The Quality Management System Approach

Clinical and Laboratory Standards Institute (CLSI) subscribes to a quality management system approach in the development of standards and guidelines, which facilitates project management; defines a document structure via a template; and provides a process to identify needed documents. The quality management system 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

C54-A-IR addresses the QSE indicated by an "X." For a description of the other documents listed in the grid, please refer to the Related CLSI Reference Materials section on the following page.

Organization	Customer Focus	Facilities and Safety	Personnel	Purchasing and Inventory	Equipment	Process Management	Documents and Records	Information Management	Nonconforming Event Management	Assessments	Continual Improvement
			C30			X C24 C30 C37 EP05 EP09 EP15 X05	C30				

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.

C54-A-IR addresses the clinical laboratory path of workflow steps indicated by an "X." For a description of the other document listed in the grid, please refer to the Related CLSI Reference Materials section on the following page.

Preexamination				**Examination**			**Postexamination**	
Examination ordering	Sample collection	Sample transport	Sample receipt/processing	Examination	Results review and follow-up	Interpretation	Results reporting and archiving	Sample management
X	C30	X			X C30		C30	

Related CLSI Reference Materials*

C24-A3 **Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions; Approved Guideline—Third Edition (2006).** This guideline provides definitions of analytical intervals, planning of quality control procedures, and guidance for quality control applications.

C30-A2 **Point-of-Care Blood Glucose Testing in Acute and Chronic Care Facilities; Approved Guideline—Second Edition (2002).** This document contains guidelines for performance of point-of-care (POC) blood glucose testing that stress quality control, training, and administrative responsibility.

C37-A **Preparation and Validation of Commutable Frozen Human Serum Pools as Secondary Reference Materials for Cholesterol Measurement Procedures; Approved Guideline (1999).** This guideline details procedures for the manufacture and evaluation of human serum pools for cholesterol measurement.

EP05-A2 **Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition (2004).** This document provides guidance for designing an experiment to evaluate the precision performance of quantitative measurement methods; recommendations on comparing the resulting precision estimates with manufacturers' precision performance claims and determining when such comparisons are valid; as well as manufacturers' guidelines for establishing claims.

EP09-A2 **Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition (Interim Revision) (2010).** This document addresses procedures for determining the bias between two clinical methods, and the design of a method comparison experiment using split patient samples and data analysis.

EP15-A2 **User Verification of Performance for Precision and Trueness; Approved Guideline—Second Edition (2006).** This document describes the demonstration of method precision and trueness for clinical laboratory quantitative methods utilizing a protocol designed to be completed within five working days or less.

M29-A3 **Protection of Laboratory Workers From Occupationally Acquired Infections; Approved Guideline—Third Edition (2005).** 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.

X05-R **Metrological Traceability and Its Implementation; A Report (2006).** This document provides guidance to manufacturers for establishing and reporting metrological traceability. A CLSI-IFCC joint project.

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

Active Membership
(As of 1 April 2008)

Sustaining Members

Abbott
American Association for Clinical Chemistry
AstraZeneca Pharmaceuticals
Bayer Corporation
BD
Beckman Coulter, Inc.
bioMérieux, Inc.
CLMA
College of American Pathologists
GlaxoSmithKline
Ortho-Clinical Diagnostics, Inc.
Pfizer Inc
Roche Diagnostics, Inc.

Professional Members

American Academy of Family Physicians
American Association for Clinical Chemistry
American Association for Laboratory Accreditation
American Association for Respiratory Care
American Chemical Society
American College of Medical Genetics
American Medical Technologists
American Society for Clinical Laboratory Science
American Society for Microbiology
American Type Culture Collection
ASCP
Associazione Microbiologi Clinici Italiani (AMCLI)
Canadian Society for Medical Laboratory Science
COLA
College of American Pathologists
College of Medical Laboratory Technologists of Ontario
College of Physicians and Surgeons of Saskatchewan
Elkin Simson Consulting Services
ESCMID
Family Health International
Hong Kong Accreditation Service Innovation and Technology Commission
International Federation of Biomedical Laboratory Science
International Federation of Clinical Chemistry
Italian Society of Clinical Biochemistry and Clinical Molecular Biology
JCCLS
The Joint Commission
National Society for Histotechnology, Inc.
Ontario Medical Association Quality Management Program-Laboratory Service
RCPA Quality Assurance Programs PTY Limited
Serbian Society of Microbiology
SIMeL
Sociedad Espanola de Bioquimica Clinica y Patologia Molecular
Sociedade Brasileira de Analises Clinicas
Sociedade Brasileira de Patologia Clinica
Turkish Society of Microbiology
Washington G2 Reports
World Health Organization

Government Members

Armed Forces Institute of Pathology
Association of Public Health Laboratories
BC Centre for Disease Control
Centers for Disease Control and Prevention
Centers for Disease Control and Prevention - Namibia
Centers for Disease Control and Prevention – Nigeria
Centers for Disease Control and Prevention – Tanzania
Centers for Medicare & Medicaid Services
Centers for Medicare & Medicaid Services/CLIA Program
Chinese Committee for Clinical Laboratory Standards
Department of Veterans Affairs
DFS/CLIA Certification
FDA Center for Biologics Evaluation and Research
FDA Center for Devices and Radiological Health
FDA Center for Veterinary Medicine
Health Canada
Massachusetts Department of Public Health Laboratories
Ministry of Health and Social Welfare – Tanzania
National Center of Infectious and Parasitic Diseases (Bulgaria)
National Health Laboratory Service (South Africa)
National Institute of Standards and Technology
National Pathology Accreditation Advisory Council (Australia)
New York State Department of Health
Ontario Ministry of Health
Pennsylvania Dept. of Health
Saskatchewan Health-Provincial Laboratory
Scientific Institute of Public Health
University of Iowa, Hygienic Lab

Industry Members

3M Medical Division
AB Biodisk
Abbott
Abbott Diabetes Care
Abbott Molecular Inc.
Abbott Point of Care Inc.
Access Genetics
Acupath
AdvaMed
Advancis Pharmaceutical Corporation
Advantage Bio Consultants, Inc.
Affymetrix, Inc. (Santa Clara, CA)
Affymetrix, Inc. (W. Sacramento, CA)
Agilent Technologies/Molecular Diagnostics
Ammirati Regulatory Consulting
Anapharm, Inc.
Anna Longwell, PC
Aptium Oncology
Arpida Ltd.
A/S Rosco
Associate Regional & University Pathologists
Astellas Pharma
AstraZeneca Pharmaceuticals
Aviir, Inc.
Axis-Shield PoC AS
Bayer Corporation – West Haven, CT
Bayer HealthCare, LLC, Diagnostics Div. – Elkhart, IN
BD
BD Biosciences – San Jose, CA
BD Diagnostic Systems
BD Vacutainer Systems
Beckman Coulter, Inc.
Beth Goldstein Consultant (PA)
Bioanalyse, Ltd.
Bio-Development S.r.l.
Biomedia Laboratories SDN BHD
bioMérieux, Inc. (MO)
bioMérieux, Inc. (NC)
Bio-Rad Laboratories, Inc. – France
Bio-Rad Laboratories, Inc. – Irvine, CA
Bio-Rad Laboratories, Inc. – Plano, TX
Blaine Healthcare Associates, Inc.
Braun Biosystems, Inc.
Canon U.S. Life Sciences, Inc.
Cempra Pharmaceuticals, Inc.
Center for Measurement Standards/ITRI
Centers for Disease Control and Prevention
Central States Research Centre, Inc.
Cepheid
Chen & Chen, LLC (IQUUM)
The Clinical Microbiology Institute
Comprehensive Cytometric Consulting
Control Lab
Copan Diagnostics Inc.
Cosmetic Ingredient Review
Cubist Pharmaceuticals
Cumbre Inc.
Dade Behring Marburg GmbH – A Siemens Company
Dahl-Chase Pathology Associates PA
David G. Rhoads Associates, Inc.
Diagnostic Products Corporation
Diagnostica Stago
Docro, Inc.
DX Tech
Eiken Chemical Company, Ltd.
Elanco Animal Health
Emisphere Technologies, Inc.
Eurofins Medinet
Fio
Focus Diagnostics
Future Diagnostics B.V.
Genomic Health, Inc.
Gen-Probe
Genzyme Diagnostics
GlaxoSmithKline
GlucoTec, Inc.
GR Micro LTD
Greiner Bio-One Inc.
Habig Regulatory Consulting
HistoGenex N.V.
Icon Laboratories, Inc.
Immunicon Corporation
Indiana State Department of Health Instrumentation Laboratory
Japan Assn. of Clinical Reagents Industries
Joanneum Research Forschungsgesellschaft mbH
Johnson & Johnson Pharmaceutical Research and Development, L.L.C.
Kaiser Permanente
K.C.J. Enterprises
Krouwer Consulting
Laboratory Specialists, Inc.
LifeScan, Inc. (a Johnson & Johnson Company)
LipoScience
Maine Standards Company, LLC
Medical Device Consultants, Inc.
Merck & Company, Inc.
Micromyx, LLC
MicroPhage
Monogen, Inc.
MultiPhase Solutions, Inc.
Nanogen
Nanogen, Point-of-Care Diagnostics Div.
Nanosphere, Inc.
Nihon Koden Corporation
Nissui Pharmaceutical Co., Ltd.
NJK & Associates, Inc.
NorDx – Scarborough Campus
NovaBiotics (Aberdeen, UK)
Novartis Institutes for Biomedical Research
Nucryst Pharmaceuticals
Olympus America, Inc.
Opti Scan Bio Medical Assoc.
Optimer Pharmaceuticals, Inc.
Orion Genomics, LLC
Ortho-Clinical Diagnostics, Inc. (Rochester, NY)
Ortho-McNeil, Inc.
Oxonica (UK)
Panaceapharma Pharmaceuticals
Paratek Pharmaceuticals, Inc.
PathCare
Pathwork Diagnostics
Pfizer Animal Health
Pfizer Inc
Phadia AB
PlaCor, Inc
Powers Consulting Services
PPD
ProSource Consulting, Inc.
QSE Consulting
Qualtek Clinical Laboratories
Quest Diagnostics, Incorporated
Radiometer America, Inc.
RCC CIDA S. A.
Replidyne
Rib-X Pharmaceuticals
Roche Diagnostics GmbH
Roche Diagnostics, Inc.
Roche Diagnostics Ltd
Roche Diagnostics Shanghai Ltd.
Roche Molecular Systems
SAIC Frederick Inc. NCI-Frederick Cancer Research & Development Center
Sanofi Pasteur
Sarstedt, Inc.
Schering Corporation
Sequenom, Inc.
Siemens Healthcare Diagnostics
Siemens Medical Solutions Diagnostics (CA)
Siemens Medical Solutions Diagnostics (DE)
Siemens Medical Solutions Diagnostics (NY)
Specialty Ranbaxy Ltd
Sphere Medical Holding Limited
State of Alabama
Stirling Medical Innovations
Streck Laboratories, Inc.
Sysmex America, Inc. (Mundelein, IL)
Sysmex Corporation (Japan)
Targanta Therapeutics, Inc
Tethys Bioscience, Inc.
TheraDoc
Therapeutic Monitoring Services, LLC
Theravance Inc.
Third Wave Technologies, Inc.
Thrombodyne, Inc.
ThromboVision, Inc.
Transasia Bio-Medicals Limited
Trek Diagnostic Systems
Upside Endeavors, LLC
Vital Diagnostics S.r.l.
Watin-Biolife Diagnostics and Medicals
Wellstat Diagnostics, LLC
Wyeth Research
XDX, Inc.
YD Consultant
ZIUR Ltd.

Trade Associations

AdvaMed
Japan Association of Clinical Reagents Industries (Tokyo, Japan)

Associate Active Members

3rd Medical Group (AK)
5th Medical Group/SGSL (ND)
22 MDSS (KS)
36th Medical Group/SGSL (Guam)
48th Medical Group/MDSS (APO, AE)
55th Medical Group/SGSAL (NE)
59th MDW/859th MDTS/MTL Wilford Hall Medical Center (TX)
Academisch Ziekenhuis-VUB
Acadiana Medical Labs, Ltd
ACL Laboratories (IL)
Adams County Hospital (OH)
Air Force Institute for Operational Health (TX)
Akron's Children's Hospital (OH)
Alameda County Medical Center
Albany Medical Center Hospital (NY)
Albemarle Hospital (NC)
Alfred I. du Pont Hospital for Children
All Children's Hospital (FL)
Allegheny General Hospital (PA)
Alpena General Hospital (MI)
Alta Bates Summit Medical Center (CA)
American University of Beirut Medical Center (NJ)
Anne Arundel Medical Center (MD)
Antelope Valley Hospital District (CA)
Arkansas Children's Hospital (AR)
Arkansas Dept of Health Public Health Laboratory (AR)
Arkansas Methodist Medical Center (AR)
Asan Medical Center (Seoul)
Asante Health System (OR)
Asiri Group of Hospitals Ltd.
Asociacion Espanola Primera de Socorros Mutuos (Uruguay)
Aspirus Wausau Hospital (WI)
Atlantic City Medical Center (NJ)
Atlantic Health Sciences Corp.
Auburn Regional Medical Center (WA)
Augusta Medical Center (VA)
Aultman Hospital (OH)
Avera McKennan (SD)
Az Sint-Jan
Azienda Ospedale Di Lecco (Italy)
Baffin Regional Hospital (Canada)
Baptist Hospital for Women (TN)
Baptist Hospital of Miami (FL)
Bassett Army Community Hospital (AK)
Baton Rouge General (LA)
Baxter Regional Medical Center (AR)
Bay Regional Medical Center (MI)
BayCare Health System (FL)
Baylor Health Care System (TX)
Bayou Pathology, APMC (LA)
Baystate Medical Center (MA)
B.B.A.G. Ve U. AS., Duzen Laboratories (Turkey)
Beebe Medical Center (DE)
Belfast HSS Trust
Beloit Memorial Hospital (WI)
Ben Taub General Hospital (TX)
The Bermuda Hospitals Board
Bonnyville Health Center (Canada)
Boston Medical Center (MA)
Boulder Community Hospital (CO)
Brantford General Hospital (Canada)
Bridgeport Hospital (CT)
Bronson Methodist Hospital (MI)
Broward General Medical Center (FL)
Calgary Laboratory Services (Calgary, AB, Canada)
California Pacific Medical Center (CA)
Cambridge Health Alliance (MA)
Camden Clark Memorial Hospital (WV)
Canadian Science Center for Human and Animal Health (Canada)
Cape Breton Healthcare Complex (Canada)
Cape Cod Hospital (MA)
Cape Fear Valley Medical Center Laboratory (NC)
Capital Health/QE II Health Sciences Centre (Nova Scotia)
Capital Health - Regional Laboratory Services (Canada)

Capital Health System Mercer Campus (NJ)
Carilion Labs Charlotte
Carpermor S.A. de C.V. (Mexico)
Catholic Health Initiatives (KY)
Cavan General Hospital (Ireland)
CDC/HIV (APO, AP)
Cedars-Sinai Medical Center (CA)
Central Baptist Hospital (KY)
Central Kansas Medical Center (KS)
Central Texas Veterans Health Care System (TX)
Centralized Laboratory Services (NY)
Centre Hospitalier Anna-Laberge (Canada)
Centura – Villa Pueblo (CO)
Chaleur Regional Hospital (Canada)
Chang Gung Memorial Hospital (Taiwan)
Changhua Christian Hospital (Taiwan)
The Charlotte Hungerford Hospital (CT)
Chatham - Kent Health Alliance (Canada)
Chesapeake General Hospital (VA)
Chester County Hospital (PA)
Children's Healthcare of Atlanta (GA)
The Children's Hospital (CO)
Children's Hospital (OH)
Children's Hospital and Medical Center (WA)
Children's Hospital & Research Center at Oakland (CA)
Children's Hospital Medical Center (OH)
Children's Hospital of Philadelphia (PA)
Children's Hospitals and Clinics (MN)
Children's Medical Center (OH)
Children's Medical Center (TX)
Children's Memorial Hospital (IL)
The Children's Mercy Hospital (MO)
Childrens Hosp. – Kings Daughters (VA)
Childrens Hospital Los Angeles (CA)
Childrens Hospital of Wisconsin (WI)
Chilton Memorial Hospital (NJ)
Christiana Care Health Services (DE)
Christus St. John Hospital (TX)
CHU Sainte – Justine (Quebec)
City of Hope National Medical Center (CA)
Clarian Health – Clarian Pathology Laboratory (IN)
Cleveland Clinic Health System Eastern Region (OH)
Clinical Labs of Hawaii (HI)
CLSI Laboratories, Univ. Pittsburgh Med. Ctr. (PA)
Colchester East Hants Health Authority (Canada)
Commonwealth of Virginia (DCLS) (VA)
Community Hospital (IN)
The Community Hospital (OH)
Community Hospital of the Monterey Peninsula (CA)
Community Medical Center (NJ)
Community Memorial Hospital (WI)
Consultants Laboratory of WI LLC (WI)
Contra Costa Regional Medical Center (CA)
Cook Children's Medical Center (TX)
Cork University Hospital (Ireland)
Corpus Christi Medical Center (TX)
Covance CLS (IN)
Covance Evansville (IN)
The Credit Valley Hospital (Canada)
Creighton Medical Laboratories (NE)
Creighton University Medical Center (NE)
Crozer-Chester Medical Center (PA)
Darwin Library NT Territory Health Services (Australia)
David Grant Medical Center (CA)
Daviess Community Hospital (IN)
Deaconess Hospital Laboratory (IN)
Deaconess Medical Center (WA)
Dean Medical Center (WI)
DeWitt Healthcare Network (USA Meddac) (VA)
DHHS NC State Lab of Public Health (NC)
Diagnostic Laboratory Services, Inc. (HI)
Diagnostic Services of Manitoba (Canada)
Diagnósticos da América S/A (Sao Paulo)
DIANON Systems/Lab Corp. (OK)
Diaz Gill-Medicina Laboratorial S.A.
Dimensions Healthcare System (MD)
Dr. Erfan & Bagedo General Hospital (Saudi Arabia)
Dr. Everette Chalmers Regional Hospital (NB)
DRAKE Center (OH)
Driscoll Children's Hospital (TX)
DSI of Bucks County (PA)
DUHS Clinical Laboratories (NC)
Dundy County Hospital (NE)
Durham VA Medical Center (NC)
DVA Laboratory Services (FL)
Dwight D. Eisenhower Medical Center (KS)
E. A. Conway Medical Center (LA)
East Central Health (Canada)
East Georgia Regional Medical Center (GA)
Eastern Health Pathology (Australia)
Easton Hospital (PA)
Edward Hospital (IL)
Effingham Hospital (GA)
Eliza Coffee Memorial Hospital (AL)
Emory University Hospital (GA)
Evangelical Community Hospital (PA)
Evans Army Community Hospital (CO)
Exeter Hospital (NH)
Federal Medical Center (MN)
First Health of the Carolinas Moore Regional Hospital (NC)
Flaget Memorial Hospital (KY)
Fletcher Allen Health Care (VT)
Fleury S.A. (Brazil)
Florida Hospital (FL)
Florida Hospital Waterman (FL)
Foote Hospital (MI)
Fort St. John General Hospital (Canada)
Forum Health Northside Medical Center (OH)
Fox Chase Cancer Center (PA)
Frankford Hospital (PA)
Fraser Health Authority Royal Columbian Hospital Site (Canada)
Fresenius Medical Care/Spectra East (NJ)
Fundacio Joan Costa Roma Consorci Sanitari de Terrassa (Spain)
Gamma-Dynacare Laboratories (Canada)
Gamma Dynacare Medical Laboratories (Ontario, Canada)
Garden City Hospital (MI)
Garfield Medical Center (CA)
Geisinger Medical Center (Danville, PA)
Genesis Healthcare System (OH)
George Washington University Hospital (DC)
Ghent University Hospital (Belgium)
Good Samaritan Hospital (OH)
Good Shepherd Medical Center (TX)
Grana S.A. (Mexico)
Grand Strand Reg. Medical Center (SC)
Gundersen Lutheran Medical Center (WI)
Guthrie Clinic Laboratories (PA)
Haga Teaching Hospital (Netherlands)
Hagerstown Medical Laboratory (MD)
Halton Healthcare Services (Canada)
Hamad Medical Corporation (Qatar)
Hamilton Regional Laboratory Medicine Program (Canada)
Hanover General Hospital (PA)
Harford Memorial Hospital (MD)
Harris Methodist Fort Worth (TX)
Health Network Lab (PA)
Health Partners Laboratories Bon Secours Richmond (VA)
Health Sciences Research Institute (Japan)
Health Waikato (New Zealand)
Heartland Health (MO)
Heidelberg Army Hospital (APO, AE)
Helen Hayes Hospital (NY)
Hema-Quebec (Canada)
Hennepin Faculty Association (MN)
Henry Ford Hospital (MI)
Henry M. Jackson Foundation (MD)
Henry Medical Center, Inc. (GA)
Hi-Desert Medical Center (CA)
Hoag Memorial Hospital Presbyterian (CA)
Holy Cross Hospital (MD)
Holy Family Medical Center (WI)
Holy Name Hospital (NJ)
Holy Spirit Hospital (PA)
Hopital Cite de La Sante de Laval (Canada)
Hopital du Haut-Richelieu (Canada)
Hôpital Maisonneuve - Rosemont (Montreal, Canada)
Hôpital Sacré-Coeur de Montreal (Quebec, Canada)
Hopital Santa Cabrini Ospedale (Canada)
Hospital Albert Einstein (Brazil)
Hospital das Clinicas-FMUSP (Brazil)
Hospital de Dirino Espirito Santa (Portugal)
The Hospital for Sick Children (Canada)
Hôtel Dieu Grace Hospital Library (Windsor, ON, Canada)
Hunter Area Pathology Service (Australia)
Imelda Hospital (Belgium)
Indiana University – Chlamydia Laboratory (IN)
Inova Fairfax Hospital (VA)
Institut fur Stand. und Dok. im Med. Lab. (Germany)
Institut National de Santé Publique du Quebec Centre de Doc. – INSPQ (Canada)
Institute Health Laboratories (PR)
Institute of Clinical Pathology and Medical Research (Australia)
Institute of Laboratory Medicine Landspitali Univ. Hospital (Iceland)
Institute of Medical & Veterinary Science (Australia)
Integrated Regional Laboratories South Florida (FL)
International Health Management Associates, Inc. (IL)
Ireland Army Community Hospital (KY)
IWK Health Centre (Canada)
Jackson County Memorial Hospital (OK)
Jackson Health System (FL)
Jackson Purchase Medical Center (KY)
Jacobi Medical Center (NY)
John C. Lincoln Hospital (AZ)
John Muir Medical Center (CA)
John T. Mather Memorial Hospital (NY)
Johns Hopkins Medical Institutions (MD)
Johns Hopkins University (MD)
Johnson City Medical Center (TN)
JPS Health Network (TX)
Kadlec Medical Center (WA)
Kaiser Permanente (CA)
Kaiser Permanente (MD)
Kaiser Permanente (OH)
Kaiser Permanente Medical Care (CA)
Kantonsspital Aarau AG (Switzerland)
Keller Army Community Hospital (NY)
Kenora-Rainy River Reg. Lab. Program (Canada)
King Fahad National Guard Health Affairs King Abdulaziz Medical City (Saudi Arabia)
King Faisal Specialist Hospital (MD)
King Hussein Cancer Center
Kings County Hospital Center (NY)
Kingston General Hospital (Canada)
Lab Medico Santa Luzia LTDA (Brazil)
Labette Health (KS)
Laboratory Alliance of Central New York (NY)
LabPlus Auckland Healthcare Services Limited (New Zealand)
Labway Clinical Laboratory Ltd (China)
Lafayette General Medical Center (LA)
Lakeland Regional Laboratories (MI)
Lakeland Regional Medical Center (FL)
Lancaster General Hospital (PA)
Landstuhl Regional Medical Center
Langley Air Force Base (VA)
LeBonheur Children's Medical Center (TN)
Legacy Laboratory Services (OR)
Lethbridge Regional Hospital (Canada)
Lewis-Gale Medical Center (VA)
L'Hotel-Dieu de Quebec (Quebec, Canada)
Licking Memorial Hospital (OH)
LifeBridge Health Sinai Hospital (MD)
LifeLabs (Canada)
Loma Linda University Medical (CA)
Long Beach Memorial Medical Center (CA)
Los Angeles County Public Health Lab. (CA)
Louisiana Office of Public Health Laboratory (LA)
Louisiana State University Medical Ctr. (LA)
Lourdes Hospital (KY)
Maccabi Medical Care and Health Fund
Madison Parish Hospital (LA)
Mafraq Hospital
Magnolia Regional Health Center (MS)
Main Line Clinical Laboratories, Inc. (PA)
Maricopa Integrated Health System (AZ)
Marquette General Hospital (MI)
Marshfield Clinic (WI)
Martha Jefferson Hospital (VA)
Martin Luther King, Jr. Harbor Hospital (CA)
Martin Memorial Health Systems (FL)
Mary Imogene Bassett Hospital (NY)
Marymount Medical Center (KY)
Massachusetts General Hospital (MA)
Massachusetts General Hospital Division of Laboratory Medicine (MA)
Maxwell Air Force Base (AL)
Mayo Clinic (MN)
Mayo Clinic Scottsdale (AZ)
MDS Metro Laboratory Services (BC, Canada)
Meadows Regional Medical Center (GA)
Mease Countryside Hospital (FL)
Medecin Microbiologiste (Canada)
Medical Center Hospital (TX)
Medical Center of Louisiana at NO-Charity (LA)
Medical Center of McKinney (TX)
Medical Centre Ljubljana (Slovenia)
Medical College of Virginia Hospital (VA)
Medical Specialists (IN)
Medical Univ. of South Carolina (SC)
MediCorp - Mary Washington Hospital (VA)
Memorial Hermann Healthcare System (TX)
Memorial Hospital at Gulfport (MS)
Memorial Hospital Laboratory (CO)
Memorial Medical Center (IL)
Memorial Medical Center (PA)
Memorial Regional Hospital (FL)
Mercy Franciscan Mt. Airy (OH)
Mercy Hospital (ME)
Mercy Medical Center (CO)
Mercy Medical Center (OR)
Methodist Hospital (MN)
Methodist Hospital (TX)
Methodist Hospital Pathology (NE)
MetroHealth Medical Center (OH)
Metropolitan Hospital Center (NY)
Metropolitan Medical Laboratory, PLC (IA)
The Michener Inst. for Applied Health Sciences (Canada)
Middelheim General Hospital
Middletown Regional Hospital (OH)
Mike O'Callaghan Federal Hospital (NV)
Mississippi Baptist Medical Center (MS)
Mississippi Public Health Lab (MS)
Monmouth Medical Center (NJ)
Montefiore Medical Center (NY)
Montreal General Hospital (Quebec)
Morton Plant Hospital (FL)
Mt. Sinai Hospital - New York (NY)
Nassau County Medical Center (NY)
National Cancer Center (S. Korea)
National Cancer Institute (MD)
National Healthcare Group (Singapore)
National Institutes of Health, Clinical Center (MD)
National Naval Medical Center (MD)
National University Hospital Department of Laboratory Medicine (Singapore)
Naval Hospital Great Lakes (IL)
Naval Hospital Oak Harbor (WA)
Naval Medical Center Portsmouth (VA)
NB Department of Health
The Nebraska Medical Center (NE)
New England Baptist Hospital (MA)
New England Fertility Institute (CT)
New Lexington Clinic (KY)
New York City Department of Health and Mental Hygiene (NY)
New York-Presbyterian Hospital (NY)
New York University Medical Center (NY)
Newark Beth Israel Medical Center (NJ)
Newton Memorial Hospital (NJ)
North Bay Hospital (FL)
North Carolina Baptist Hospital (NC)
North Coast Clinical Laboratory, Inc. (OH)
North District Hospital (Hong Kong)
North Mississippi Medical Center (MS)
North Shore-Long Island Jewish Health System Laboratories (NY)
Northeast Pathologists, Inc. (MO)
Northridge Hospital Medical Center (CA)
Northside Hospital (GA)
Northwest Texas Hospital (TX)
Northwestern Memorial Hospital (IL)
Norton Healthcare (KY)
Ochsner Clinic Foundation (LA)
Ohio State University Hospitals (OH)
Onze Lieve Vrouw Ziekenhuis (Belgium)
Ordre Professionel des Technologistes Medicaux du Quebec (Quebec)
Orebro University Hospital
Orlando Regional Healthcare System (FL)
The Ottawa Hospital (Canada)
Our Lady of Lourdes Medical Center (NJ)
Our Lady of Lourdes Reg. Medical Ctr. (LA)
Our Lady of the Way Hospital (KY)
Our Lady's Hospital for Sick Children (Ireland)
Overlake Hospital Medical Center (WA)
Palmetto Health Baptist Laboratory (SC)
Pathologists Associated (IN)
Pathology and Cytology Laboratories, Inc. (KY)
Pathology Associates Medical Laboratories (WA)
Penn State Hershey Medical Center (PA)
Pennsylvania Hospital (PA)
Penrose St. Francis Health Services (CO)
The Permanente Medical Group (CA)
Perry County Memorial Hospital (IN)
Peterborough Regional Health Centre (Canada)
Piedmont Hospital (GA)
Pitt County Memorial Hospital (NC)
Prairie Lakes Hospital (SD)
Presbyterian Hospital of Dallas (TX)
Presbyterian/St. Luke's Medical Center (CO)
Prince County Hospital

Princess Margaret Hospital (Hong Kong)
Providence Alaska Medical Center (AK)
Providence Health Care (Canada)
Providence Medford Medical Center (OR)
Provincial Health Services Authority (Vancouver, BC, Canada)
Provincial Laboratory for Public Health (Edmonton, AB, Canada)
Queen Elizabeth Hospital (Canada)
Queensland Health Pathology Services (Australia)
Quest Diagnostics, Inc
Quest Diagnostics, Inc (San Juan Capistrano, CA)
Quest Diagnostics JV (IN, OH, PA)
Quest Diagnostics Laboratories (WA)
Quincy Hospital (MA)
Rady Children's Hospital San Diego (CA)
Redington-Fairview General Hospital (ME)
Regional Health Authority Four (RHA4) (Canada)
Regions Hospital (MN)
Reid Hospital & Health Care Services (IN)
Renown Regional Medical Center (NV)
Research Medical Center (MO)
Riverside Regional Medical Center (VA)
Riyadh Armed Forces Hospital, Sulaymainia
Robert Wood Johnson University Hospital (NJ)
Roxborough Memorial Hospital (PA)
Royal Victoria Hospital (Canada)
Rush North Shore Medical Center (IL)
SAAD Specialist Hospital (Saudi Arabia)
Sacred Heart Hospital (FL)
Sacred Heart Hospital (WI)
Sahlgrenska Universitetssjukhuset (Sweden)
Saint Elizabeth Regional Medical Center (NE)
Saint Francis Hospital & Medical Center (CT)
Saint Mary's Regional Medical Center (NV)
Saints Memorial Medical Center (MA)
St. Agnes Healthcare (MD)
St. Anthony Hospital (OK)
St. Anthony Hospital Central Laboratory (CO)
St. Anthony's Hospital (FL)
St. Barnabas Medical Center (NJ)
St. Christopher's Hospital for Children (PA)
St. Elizabeth Community Hospital (CA)
St. Francis Hospital (SC)
St. Francis Medical Center (MN)
St. John Hospital and Medical Center (MI)
St. John's Hospital (IL)
St. John's Hospital & Health Ctr. (CA)
St. John's Mercy Medical Center (MO)
St. John's Regional Health Center (MO)
St. Joseph Medical Center (MD)
St. Joseph Mercy – Oakland (MI)
St. Joseph Mercy Hospital (MI)
St. Joseph's Hospital (FL)
St. Joseph's Hospital & Health Center (ND)
St. Joseph's Medical Center (GA)
St. Joseph's Regional Medical Center (NJ)
St. Jude Children's Research Hospital (TN)
St. Louis University Hospital (MO)
St. Luke's Hospital (FL)
St. Luke's Hospital (IA)
St. Luke's Hospital (PA)
St. Margaret Memorial Hospital (PA)
St. Martha's Regional Hospital (Canada)
St. Mary Medical Center (CA)
St. Mary's Health Center (MO)
St Mary's Healthcare (SD)
St. Mary's Medical Center (IN)
St. Michael's Hospital Diagnostic Laboratories & Pathology (Canada)
St. Tammany Parish Hospital (LA)
St. Thomas More Hospital (CO)
Sampson Regional Medical Center (NC)
Samsung Medical Center (Korea)
San Francisco General Hospital-University of California San Francisco (CA)
Sanford USP Medical Center (SD)
SARL Laboratoire Carron (France)
Saudi Aramco Medical (Saudi Arabia)
Scott Air Force Base (IL)
Scott & White Memorial Hospital (TX)
Seoul National University Hospital (Korea)
Seton Medical Center (CA)
Shamokin Area Community Hospital (PA)
Sheik Kalifa Medical City (UAE)
Shore Memorial Hospital (NJ)
Shriners Hospitals for Children (SC)
Singapore General Hospital (Singapore)
SJRMC Plymouth Laboratory (IN)
Sky Lakes Medical Center (OR)
South Bend Medical Foundation (IN)
South County Hospital (RI)
South Dakota State Health Laboratory (SD)
South Miami Hospital (FL)
Southern Health Care Network (Australia)
Southern Maine Medical Center (ME)
Southwest Nova District Health Authority (Canada)
Speare Memorial Hospital (NH)
Spectrum Health - Blodgett Campus (MI)
Stanford Hospital and Clinics (CA)
State of Connecticut Department of Public Health (CT)
State of Hawaii Department of Health (HI)
State of Washington-Public Health Labs (WA)
Steele Memorial Hospital (ID)
Stillwater Medical Center (OK)
Stony Brook University Hospital (NY)
Stormont-Vail Regional Medical Center (KS)
Sudbury Regional Hospital (Canada)
Suncoast Medical Clinic (FL)
Sunnybrook Health Science Center (ON, Canada)
Sunrise Hospital and Medical Center (NV)
Swedish Medical Center (CO)
Sydney South West Pathology Service (Australia)
T.J. Samson Community Hospital (KY)
Taipei Veterans General Hospital (Taiwan)
Taiwan Society of Laboratory Medicine
Tallaght Hospital
Tartu University Clinics (Tartu)
Temple Univ. Hospital - Parkinson Pav. (PA)
Texas Children's Hospital (TX)
Texas Department of State Health Services
Thomason Hospital (TX)
Timmins and District Hospital (Canada)
The Toledo Hospital (OH)
Touro Infirmary (LA)
Tri-Cities Laboratory (WA)
Trident Medical Center (SC)
Trinity Medical Center (AL)
Tripler Army Medical Center (HI)
Tufts New England Medical Center Hospital (MA)
Tulane Medical Center Hospital & Clinic (LA)
Turku University Central Hospital
UCI Medical Center (CA)
UCLA Medical Center Clinical Laboratories (CA)
UCSD Medical Center (CA)
UCSF Medical Center China Basin (CA)
UMass Memorial Medical Center (MA)
UMC of Southern Nevada (NV)
UNC Hospitals (NC)
Union Clinical Laboratory (Taiwan)
United Christian Hospital (Hong Kong)
United Clinical Laboratories (IA)
Unity HealthCare (IA)
Universita Campus Bio-Medico (Italy)
Universitair Ziekenhuis Antwerpen (Belgium)
University College Hospital (Ireland)
University Medical Center at Princeton (NJ)
University of Alabama-Birmingham Hospital (AL)
University of Arkansas for Medical Sci. (AR)
University of Chicago Hospitals (IL)
University of Colorado Health Sciences Center (CO)
University of Colorado Hospital
University of Iowa Hospitals and Clinics (IA)
University of Kentucky Med. Ctr. (KY)
University of Maryland Medical System (MD)
University of Medicine & Dentistry, NJ University Hosp. (NJ)
University of Miami (FL)
University of Missouri Hospital (MO)
University of MN Medical Center - Fairview
University of MS Medical Center (MS)
University of So. Alabama Children's and Women's Hospital (AL)
University of Texas Health Center (TX)
The University of Texas Medical Branch (TX)
University of the Ryukyus (Japan)
University of Virginia Medical Center
University of Washington
UPMC Bedford Memorial (PA)
U.S.A. Meddac (Pathology Division) (MO)
UW Hospital (WI)
UZ-KUL Medical Center (Belgium)
VA (Asheville) Medical Center (NC)
VA (Bay Pines) Medical Center (FL)
VA (Chillicothe) Medical Center (OH)
VA (Cincinnati) Medical Center (OH)
VA (Dallas) Medical Center (TX)
VA (Dayton) Medical Center (OH)
VA (Decatur) Medical Center (GA)
VA (Hines) Medical Center (IL)
VA (Indianapolis) Medical Center (IN)
VA (Iowa City) Medical Center (IA)
VA (Long Beach) Medical Center (CA)
VA (Miami) Medical Center (FL)
VA New Jersey Health Care System (NJ)
VA Outpatient Clinic (OH)
VA (Phoenix) Medical Center (AZ)
VA (San Diego) Medical Center (CA)
VA (Seattle) Medical Center (WA)
VA (Sheridan) Medical Center (WY)
VA (Tucson) Medical Center (AZ)
Valley Health (VA)
Vancouver Hospital and Health Sciences Center (BC, Canada)
Vancouver Island Health Authority (Canada)
Vanderbilt University Medical Center (TN)
Via Christi Regional Medical Center (KS)
Virga Jessezieukenhuis (Belgium)
ViroMed Laboratories (LabCorp) (MN)
Virtua - West Jersey Hospital (NJ)
WakeMed (NC)
Walter Reed Army Medical Center (DC)
Warren Hospital (NJ)
Washington Hospital Center (DC)
Waterbury Hospital (CT)
Waterford Regional Hospital (Ireland)
Wayne Memorial Hospital (NC)
Weirton Medical Center (WV)
Wellstar Douglas Hospital Laboratory (GA)
Wellstar Paulding Hospital (GA)
Wellstar Windy Hill Hospital Laboratory (GA)
West China Second University Hospital, Sichuan University (P.R. China)
West Valley Medical Center Laboratory (ID)
Westchester Medical Center (NY)
Western Baptist Hospital (KY)
Western Healthcare Corporation (Canada)
Wheaton Franciscan & Midwest Clinical Laboratories (WI)
Wheeling Hospital (WV)
Whitehorse General Hospital (Canada)
William Beaumont Army Medical Center (TX)
William Beaumont Hospital (MI)
William Osler Health Centre (Canada)
Winchester Hospital (MA)
Winn Army Community Hospital (GA)
Wisconsin State Laboratory of Hygiene (WI)
Wishard Health Sciences (IN)
Womack Army Medical Center (NC)
Woodlawn Hospital (IN)
York Hospital (PA)

950 West Valley Road ▾ Suite 2500 ▾ Wayne, PA 19087 ▾ USA ▾ PHONE 610.688.0100 ▾ FAX 610.688.0700
customerservice@clsi.org ▾ www.clsi.org ▾ ISBN 1-56238-851-7 (Print) ▾ ISBN 1-56238-852-5 (Electronic)

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

ifcc International Federation of Clinical Chemistry and Laboratory Medicine