September 2014

EP05-A3

Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline—Third Edition

This document provides guidance for evaluating the precision performance of quantitative measurement procedures. It is intended for manufacturers of quantitative measurement procedures and for laboratories that develop or modify such procedures.

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

Clinical and Laboratory Standards Institute

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

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

Consensus Process

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

Commenting on Documents

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

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

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

Appeals Process

If it is believed that an objection has not been adequately addressed, the process for appeals is documented in the CLSI Standards Development Policies and Process document.

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

Get Involved—Volunteer!

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

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

ISBN 1-56238-967-X (Print)
ISBN 1-56238-968-8 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic)

EP05-A3
Vol. 34 No. 13
Replaces EP05-A2
Vol. 24 No. 25

Volume 34 Number 13

Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline—Third Edition

Robert J. McEnroe, PhD
A. Paul Durham
Marc D. Goldford
Marina V. Kondratovich, PhD
Samir Lababidi, PhD
Robert Magari, PhD
Jonathan Guy Middle, PhD
James F. Pierson-Perry
Jeffrey E. Vaks, PhD

Abstract

Clinical and Laboratory Standards Institute document EP05-A3—*Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline—Third Edition* provides guidance for evaluating the precision of *in vitro* diagnostic quantitative measurement experimental designs and includes recommendations for establishing precision performance. Included are guidelines for duration, experimental designs, materials, data analysis, summarization, and interpretation—techniques adaptable for a wide spectrum of measurands and system complexity. These guidelines are intended for manufacturers or developers of clinical laboratory measurement procedures, and for users who wish to determine their own performance capabilities. A balance is created in the document between complexity of design and analysis, and simplicity of operation.

Clinical and Laboratory Standards Institute (CLSI). *Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline—Third Edition*. CLSI document EP05-A3 (ISBN 1-56238-967-X [Print]; ISBN 1-56238-968-8 [Electronic]). Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087 USA, 2014.

The Clinical and Laboratory Standards Institute consensus process, which is the mechanism for moving a document through two or more levels of review by the health care community, is an ongoing process. Users should expect revised editions of any given document. Because rapid changes in technology may affect the procedures, methods, and protocols in a standard or guideline, users should replace outdated editions with the current editions of CLSI documents. Current editions are listed in the CLSI catalog and posted on our website at www.clsi.org.

If you or your organization is not a member and would like to become one, and to request a copy of the catalog, contact us at:

P: 610.688.0100 **F:** 610.688.0700 **E:** customerservice@clsi.org **W:** www.clsi.org.

Copyright ©2014 Clinical and Laboratory Standards Institute. Except as stated below, any reproduction of content from a CLSI copyrighted standard, guideline, companion product, or other material requires express written consent from CLSI. All rights reserved. Interested parties may send permission requests to permissions@clsi.org.

CLSI hereby grants permission to each individual member or purchaser to make a single reproduction of this publication for use in its laboratory procedure manual at a single site. To request permission to use this publication in any other manner, e-mail permissions@clsi.org.

Suggested Citation

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

Proposed Guideline

December 1981

Tentative Guideline

December 1982

Approved Guideline

February 1999

Approved Guideline—Second Edition

August 2004

Approved Guideline—Third Edition

September 2014

ISBN 1-56238-967-X (Print)
ISBN 1-56238-968-8 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic)

Committee Membership

Consensus Committee on Evaluation Protocols

James F. Pierson-Perry
Chairholder
Siemens Healthcare Diagnostics, Inc.
Newark, Delaware, USA

Mitchell G. Scott, PhD
Vice-Chairholder
Barnes-Jewish Hospital
St. Louis, Missouri, USA

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

Robert J. McEnroe, PhD
Roche Diagnostics, Inc.
Indianapolis, Indiana, USA

James H. Nichols, PhD, DABCC, FACB
Vanderbilt University School of Medicine
Nashville, Tennessee, USA

Gene Pennello, PhD
FDA Center for Devices and Radiological Health
Silver Spring, Maryland, USA

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

Document Development Committee on Evaluation of Precision

Robert J. McEnroe, PhD
Chairholder
Roche Diagnostics, Inc.
Indianapolis, Indiana, USA

Victor R. De Jesus, PhD
Centers for Disease Control and Prevention
Atlanta, Georgia, USA

Marc D. Goldford
r2 Diagnostics
South Bend, Indiana, USA

Walter W. Hauck, PhD
USP
Rockville, Maryland, USA

Ming Jin, PhD
University of Illinois Medical Center
Chicago, Illinois, USA

Samir Lababidi, PhD
FDA Center for Biologics Evaluation and Research
Rockville, Maryland, USA

Robert Magari, PhD
Beckman Coulter
Miami, Florida, USA

Jonathan Guy Middle, PhD
Birmingham, United Kingdom

James F. Pierson-Perry
Siemens Healthcare Diagnostics, Inc.
Newark, Delaware, USA

Staff

Clinical and Laboratory Standards Institute
Wayne, Pennsylvania, USA

Luann Ochs, MS
Senior Vice President – Operations
Staff Liaison

Ron S. Quicho, MS
Project Manager

Megan L. Tertel, MA
Editorial Manager

Joanne P. Christopher, MA
Editor

Patrice E. Polgar
Editor

Acknowledgment

CLSI and the Consensus Committee on Evaluation Protocols gratefully acknowledge the following individuals for their help in preparing this document:

A. Paul Durham
APD Consulting
Culver City, California, USA

Marina V. Kondratovich, PhD
FDA Center for Devices and Radiological Health
Silver Spring, Maryland, USA

Jeffrey E. Vaks, PhD
Roche Molecular Systems
Pleasanton, California, USA

Contents

Abstract . . . i

Committee Membership . . . iii

Foreword . . . vii

Chapter 1: Introduction . . . 1

1.1 Scope . . . 2

1.2 Standard Precautions . . . 3

1.3 Terminology . . . 3

1.4 Overview of the Precision Evaluation Process . . . 8

1.5 Introduction to Basic Concepts . . . 10

Chapter 2: Selecting and Optimizing an Appropriate Precision Evaluation Protocol . . . 27

2.1 The Need for Multiple Study Options . . . 30

2.2 Study Design Selection Considerations . . . 30

2.3 Suggested Study Designs . . . 35

2.4 Multifactor Studies . . . 35

Chapter 3: Single-Site Precision Evaluation Study . . . 37

3.1 Introduction . . . 38

3.2 General Guidelines . . . 39

3.3 Preliminary Checkout . . . 39

3.4 Protocol Design and Requirements . . . 40

3.5 Experimental Steps . . . 43

3.6 Data Analysis . . . 44

3.7 Combining Precision Study Results . . . 50

3.8 Suggested Format for Summarizing Single-Site Precision Study Results . . . 52

3.9 Note on Unbalanced Datasets . . . 53

Chapter 4: Multisite Precision Evaluation Study. 55

4.1 Overview 56
4.2 General Guidelines 57
4.3 Preliminary Checkout 57
4.4 Protocol Design and Requirements 58
4.5 Experimental Steps 61
4.6 Data Analysis 62
4.7 Suggested Format for Summarizing Multisite Precision Study Results 67
4.8 Note on Unbalanced Datasets 68

Chapter 5: Conclusion 69

Chapter 6: Supplemental Information 71

References 72
Appendix A. Worked Example—Single-Site Study 76
Appendix B. Worked Example—Multisite Study 80
Appendix C. Advanced Models 89
The Quality Management System Approach 94
Related CLSI Reference Materials 95

Foreword

Current clinical laboratory literature contains numerous examples of product evaluations. For characterizing basic precision types, many of these examples use the basic concepts, designs, and analyses discussed in this guideline. In special cases, more complex and customized experimental designs have been used for both published studies and regulatory purposes. However, there remains a strong need in the clinical laboratory community for the basic rationales and approaches described in this document for assessing the precision performance characteristics of quantitative measurement procedures.

The great diversity of *in vitro* diagnostic devices currently available makes it impossible to recommend a single experimental design for all measurement procedures and associated devices. Nevertheless, requirements for materials, procedures, data analysis, and interpretation must be adaptable for the widest possible variety of measurands and instruments. In developing the standardized protocols in this document, many recommendations for duration, inclusion of QC procedures, and methods of determining the relevant sources of variation were carefully considered. The resulting protocols represent a balance between complexity of design and data analysis, and simplicity of implementation and efficiency. This document was written to provide general guidance consistent with other international consensus standards.

KEY WORDS

Analysis of variance

Evaluation protocol

Experimental design

Imprecision

Outliers

Precision

Precision profile

Quality control

Repeatability

Reproducibility

Within-laboratory precision

Overview of Changes

The third edition narrows the scope of EP05 by limiting its discussion of single-site experimental designs to procedures suitable for establishing or validating precision performance characteristics. Accordingly, EP05 is now addressed primarily to manufacturers and developers. Recommendations for end-user laboratories for verifying repeatability and within-laboratory precision claims can be found in CLSI document EP15[1]. The precision verification protocol in that guideline has been tailored for compatibility with EP05's single-site study designs.

The single-site protocol familiar from previous editions of EP05—calling for measurements on 20 days, with two runs per day and two replicates per run for a given sample, reagent lot, etc.—is retained in this third edition as a standardized experiment for use by manufacturers and developers in evaluating the repeatability and within-laboratory (within-device) precision of a measurement procedure (or "assay").

No matter how these performance characteristics are established, it is important that the assessments be verifiable, and that they characterize precision over a substantial period of time and across most of the assay's stated measuring interval. The single-site experimental designs described in EP05 meet these requirements (see Chapter 3). It is expected that the original "20 × 2 × 2" design will continue to serve well for the great majority of quantitative assays used in clinical laboratories. However, extensive guidance was added on how to optimize this design for a given assay in light of its sources of variation and their relative magnitudes and interrelationships (see Chapter 2).

Moreover, in recognition of the wide diversity of quantitative devices in use today, which differ in character and complexity, variants of the 20 × 2 × 2 design are also discussed. Appendix C is devoted to advanced models—multifactor designs—for use when a two-factor design lacks the ability to do justice to the major sources of

Due to the complex nature of the calculations in this guideline, it is recommended that the user have access to a computer and statistical software, such as **StatisPro™ method evaluation software from CLSI.**

variation affecting an assay's within-laboratory precision. Depending on the assay, some of these models should also prove useful to manufacturers for the insights they can yield both during assay development and optimization and after the assay enters routine production.

New to EP05 is a second standardized experiment: a multisite protocol calling, minimally, for repeated measurements at three sites on five days. Both 3 (sites) × 5 (days) × 5 (replicates per day) and 3 (sites) × 5 (days) × 2 (runs per day) × 3 (replicates per run) implementations are described (see Chapter 4). This ancillary protocol addresses site-to-site variability and estimates of reproducibility. It has been tailored for suitability in the context of validating a new assay, when such a study may be required due to the assay's character and/or to regulatory demands.

To help foster understanding of basic concepts, the new edition includes an extensive tutorial for the nonstatistician (see Section 1.5). Numerical examples illustrating a single-site 20 × 2 × 2 study and a complete multisite 3 × 5 × 5 study are presented in the appendixes.

Consistency With International Standards

EP05 is largely consistent with recommendations in the ISO 5725 series of standards, particularly ISO 5725-3.[2] EP05's single-site study incorporates the basic concepts in ISO 5725-2.[3] Whereas the ISO 5725 perspective places primary emphasis on interlaboratory sources of variation, EP05 has focused on within-laboratory sources of variation accumulating over time. However, EP05's newly introduced multisite study addresses site-to-site sources of variation and estimates of reproducibility.

Furthermore, while the ISO 5725 series requires characterizing both repeatability and reproducibility across the entire measuring interval, this is encouraged (but not required) in EP05.

Chapter 1

Introduction

This chapter includes:

- Document scope and applicable exclusions
- Background information pertinent to the document content
- Standard Precautions information, as applicable
- Note on terminology
- Definitions
- Abbreviations and acronyms

Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline—Third Edition

1 Introduction

1.1 Scope

This document provides guidance for studies intended to establish the within-site precision performance characteristics of quantitative measurement procedures used in clinical laboratories, and also for studies addressing site-to-site variability. Multiple experimental protocols are described, along with considerations on how to select and optimize the protocol(s) best suited for a specific measurement procedure (or "assay") and its intended use.

1.1.1 Intended Users

Intended users of this document are:

- Developers of a new measurement procedure who wish to establish its precision characteristics, be it a manufacturer that wants to distribute the product to multiple laboratories, or a clinical laboratory developing it for their own use
- End users who modify an existing assay and therefore need to reassess its precision performance
- Users who want to understand how precision performance estimates are established and/or want to perform rigorous precision assessments of their own
- Manufacturers in need of advanced methods (see Appendix C) for obtaining deeper insights into the precision characteristics of a quantitative measurement procedure during assay development, optimization, and routine manufacture

It is assumed that readers of this document have some familiarity with statistical data analysis, including basic analysis of variance (ANOVA), or access to statistical support resources. Section 1.5 provides a brief introduction to several of the basic concepts involved; while CLSI document EP15[1] includes a detailed discussion of one-way ANOVA.

1.1.2 Limitations on Use

Those wishing only to verify a manufacturer's claims for the precision of a quantitative clinical measurement procedure should follow the guidance in CLSI document EP15.[1]

The protocols in this document may not be applicable to some quantitative measurement procedures for which appropriate test materials do not exist. In particular, the standardized single-site and multisite procedures are not directly applicable to measurement procedures involving samples with **inadequate stability** (eg, RBC count or blood gas determinations)

© Clinical and Laboratory Standards Institute. All rights reserved.

and/or **inadequate volume** (eg, fluorescence *in situ* hybridization or immunohistochemistry determinations on tissue samples).

The analytical process is assumed to be stable with no degradation of samples or reagents taking place during the period of data collection. This leads to an assumption that variation of measurement results is random, and variation coming from each source being considered is independently and normally distributed. Evaluation of reagent degradation is addressed in CLSI document EP25.[4] Evaluation of sample degradation is addressed in CLSI document C56.[5]

This document provides guidelines for use with quantitative measurement procedures and instruments, including those on which qualitative determinations are based. Evaluation of precision performance for fully qualitative measurement procedures or instruments, however, is beyond the scope of this guideline. For this type of evaluation, consult CLSI document EP12.[6]

1.2 Standard Precautions

Because it is often impossible to know what isolates or specimens might be infectious, all patient and laboratory specimens are treated as infectious and handled according to "standard precautions." Standard precautions are guidelines that combine the major features of "universal precautions and body substance isolation" practices. Standard precautions cover the transmission of all known infectious agents and thus are more comprehensive than universal precautions, which are intended to apply only to transmission of bloodborne pathogens. The Centers for Disease Control and Prevention address this topic in published guidelines that address the daily operations of diagnostic medicine in human and animal medicine while encouraging a culture of safety in the laboratory.[7] For specific precautions for preventing the laboratory transmission of all known infectious agents from laboratory instruments and materials and for recommendations for the management of exposure to all known infectious diseases, refer to CLSI document M29.[8]

1.3 Terminology

1.3.1 A Note on Terminology

CLSI, as a global leader in standardization, is firmly committed to achieving global harmonization wherever possible. Harmonization is a process of recognizing, understanding, and explaining differences while taking steps to achieve worldwide uniformity. CLSI recognizes that medical conventions in the global metrological community have evolved differently in the United States, Europe, and elsewhere; that these differences are reflected in CLSI, International Organization for Standardization (ISO), and European Committee for Standardization (CEN) documents; and that legally required use of terms, regional usage, and different consensus timelines are all

© Clinical and Laboratory Standards Institute. All rights reserved.

In EP05, the term **total precision** was replaced with **within-laboratory precision** or **within-device precision.** Either of these new preferred terms may be used.

important considerations in the harmonization process. In light of this, CLSI's consensus process for development and revision of standards and guidelines focuses on harmonization of terms to facilitate the global application of standards and guidelines.

In EP05 (beginning with the second approved edition of the guideline, EP05-A2), the term *total precision* was replaced with *within-laboratory* precision or *within-device precision*. Either of these new preferred terms may be used.

In EP05, reproducibility encompasses sources of variation associated with multiple laboratories and/or multiple instruments, in addition to those sources affecting within-laboratory precision. This differs somewhat from the ISO characterization. Within the ISO framework a "reproducibility" precision type must encompass different locations **and** different operators **and** different measuring systems (instruments). On the other hand, the ISO framework presents no requirements for time, whereas it is implicit in the definition adopted here that a period of time encompassing at least five days is required.

1.3.2 Definitions

assay – see **measurement procedure.**

balanced – describes an experimental design or dataset wherein all cells ("treatment combinations") have the same number of observations. A design or dataset is unbalanced if this condition fails to hold; **NOTE:** In concrete terms, for precision studies of the kind discussed in Chapters 3 and 4 of this guideline, the experimental design is balanced if, for any given sample, it specifies the same number of replicates for each run; the same number of runs for each day; and, for a multisite study, the same number of days for each site. Otherwise, the design is unbalanced. (Note that all designs recommended in the main document are balanced.) The corresponding dataset is balanced if it has the same number of results for each run, day, and site—ie, if the design is balanced and no results are missing or omitted from the analysis. Otherwise, the dataset is unbalanced.

crossed – in the design of a multifactor study, the effects of a factor A are said to be crossed with the effects of a factor B if some of the effects of factor A occur with more than one effect of factor B; **EXAMPLE:** If the reagent lots that are studied at each of several sites are common across the sites, then reagent lot is crossed with site. Factors A and B are fully crossed if every effect of factor A occurs with every effect of factor B.

imprecision – the random dispersion of a set of replicate measurements and/or values expressed quantitatively by a statistic, such as standard deviation or coefficient of variation.

© Clinical and Laboratory Standards Institute. All rights reserved.

intermediate precision – measurement precision under a set of intermediate precision conditions of measurement (JCGM 200:2012).[9]

intermediate precision conditions – conditions where test results or measurement results are obtained with the same method, on identical test/measurement items in the same test or measurement facility, under some different operating conditions; **NOTE 1:** There are four elements to the operating conditions: time, calibration, operator, and equipment (ISO 3534-2)[10]; **NOTE 2:** The changed elements in operating conditions must be noted; this could include precision estimates commonly called, for example "between-run," "within-day," "between-day," "within-device," and "within-laboratory."

measurand – quantity intended to be measured (JCGM 200:2012)[9]; **NOTE 1:** The term "measurand" and definition encompass all quantities, while the commonly used term "analyte" refers to a tangible entity subject to measurement. For example, "substance" concentration is a quantity that may be related to a particular analyte; **NOTE 2:** In the type of quantity "mass of protein in 24-hour urine," "protein" is the analyte. In "amount of substance of glucose in plasma," "glucose" is the analyte. In both cases, the long phrase represents the measurand (ISO 17511)[11]; **NOTE 3:** In the type of quantity "catalytic concentration of lactate dehydrogenase isoenzyme 1 in plasma," "lactate dehydrogenase isoenzyme 1" is the analyte (ISO 18153).[12]

measurement procedure – detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result (JCGM 200:2012)[9]; **NOTE 1:** This term pertains to specific procedures as marketed by specific manufacturers. **NOTE 2:** In this document, "assay" is sometimes used as a stylistic variant for "measurement procedure."

measuring interval – set of values of quantities of the same kind that can be measured by a given measuring instrument or measuring system with specified instrumental measurement uncertainty, under defined conditions (JCGM 200:2012)[9]; **NOTE 1:** In some fields, the term is "measuring range" or "measurement range" (JCGM 200:2012)[9]; **NOTE 2:** The lower limit of a measuring interval should not be confused with detection limit (JCGM 200:2012)[9]; **NOTE 3:** This represents the interval of *in vitro* diagnostic (IVD) examination results over which the performance characteristics of the IVD medical device were validated by the manufacturer; **NOTE 4:** Formerly, the terms "reportable range" or "measuring range" were used; **NOTE 5:** The measuring interval over which the performance characteristics of an IVD medical device have been validated has been called the reportable range (ISO 18113-1)[13]; **NOTE 6:** For a discussion of the difference between interval and range, see A.2.11 of ISO

© Clinical and Laboratory Standards Institute. All rights reserved.

18113-1)[13]; **NOTE 7:** In the United States, the term often used is "analytical measuring range"; **NOTE 8:** In laboratory medicine the term is defined as "measuring range," or "analytical measurement range"; **NOTE 9:** For this document, the range of values (in units appropriate for the measurand) over which the acceptability criteria for the measurement procedure have been met; ie, the intersection of the linearity interval, the interval limited by the lower and upper limits of quantitation, and the interval representing acceptable imprecision.

nested – in the design of a multifactor study, the effects of a factor A are said to be nested within the effects of a factor B if each effect of factor A occurs with only one effect of factor B; **EXAMPLE:** Replicates within a run are distinct from replicates within another run, thus replicates are nested within runs.

precision (measurement) – closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions (JCGM 200:2012)[9]; **NOTE 1:** Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement (JCGM 200:2012)[9]; **NOTE 2:** The "specified conditions" can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement (see ISO 5725-3:1994)[2] (JCGM 200:2012)[9]; **NOTE 3:** Measurement precision is used to define measurement repeatability, intermediate measurement precision, and measurement reproducibility (JCGM 200:2012).[9]

repeatability (measurement) – measurement precision under a set of repeatability conditions of measurement (JCGM 200:2012)[9]; **NOTE:** Formerly termed "within-run precision."

repeatability condition (of measurement) – condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time (JCGM 200:2012)[9]; **NOTE 1:** Conditions where independent measurement results are obtained with the same method of measurement on identical samples ("test items" ISO 15198)[14] in the same laboratory (or "location" in ISO 15198)[14] by the same operator using the same equipment within a short interval of time. For the purposes of ISO 17593,[15] "laboratories" should be interpreted as "locations" (ISO 17593)[15]; **NOTE 2:** Essentially unchanged conditions, intended to represent conditions resulting in minimum variability of test results (ISO 15198)[14]; **NOTE 3:** Repeatability can be determined from replicated measurements made under the repeatability conditions and also from differences

© Clinical and Laboratory Standards Institute. All rights reserved.

between duplicate measurements made within various runs on different days. The latter method is used with experimental designs described in this document.

reproducibility (measurement) – measurement precision under reproducibility conditions of measurement (JCGM 200:2012)[9]; **NOTE 1:** Relevant statistical terms are given in ISO 5725-1[16] and ISO 5725-2[3] (JCGM 200:2012)[9]; **NOTE 2:** More broadly defined as the agreement between repeated measurements on the same specimen under different conditions; **NOTE 3:** The closeness of the agreement between the results of measurements of the same measurand, where the measurements are carried out under changed conditions; **NOTE 4:** Changed conditions may include principle or method of measurement, observer, measuring instrument, location, conditions of use, and time.

reproducibility condition (of measurement) – condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects (JCGM 200:2012)[9]; **NOTE 1:** A specification should give the conditions changed and unchanged, to the extent practical (JCGM 200:2012)[9]; **NOTE 2:** The changed conditions may refer to different lots, runs, time (day), technicians, etc.; **NOTE 3:** In chemistry, the terms "between-laboratories," "interlaboratory," or "among-laboratories precision" condition of measurement are sometimes used to designate this concept.

unbalanced – see **balanced.**

uncertainty (of measurement) – non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used (JCGM 200:2012).[9]

1.3.3 Abbreviations and Acronyms

%CV coefficient of variation expressed as a percentage

$\%CV_R$ repeatability coefficient of variation

$\%CV_{WL}$ within-laboratory coefficient of variation

ANOVA analysis of variance

CB confidence band

CEN Comité Européen de Normalisation (European Committee for Standardization)

CI confidence interval

CL confidence level

DF, df degrees of freedom

df_R degrees of freedom for repeatability

© Clinical and Laboratory Standards Institute. All rights reserved.

df_{REP}	degrees of freedom for reproducibility
df_{WL}	degrees of freedom for within-laboratory precision
ELISA	enzyme-linked immunosorbent assay
EMS	expected value of the mean square
ISO	International Organization for Standardization
IVD	*in vitro* diagnostic
LLMI	lower limit of the measuring interval
LoD	limit of detection
LoQ	limit of quantitation
MAD	median absolute deviation
MINQUE	minimum norm quadratic unbiased estimation
MIVQUE	minimum variance quadratic unbiased estimation
MS	mean squares
PI	package insert (or equivalent document)
PT	proficiency testing
QC	quality control
REML	restricted maximum likelihood
SD	standard deviation
SS	sum of squares
ULMI	upper limit of the measuring interval
Var	variance

1.3.4 Symbols

μ	population (or "true") mean
x	sample mean
χ^2	chi-square
$\chi^2_{CL,DF}$	quantile of the chi-square distribution for confidence level and degrees of freedom

1.4 Overview of the Precision Evaluation Process

Figure 1 shows a basic overview of the process involved in evaluating the precision of quantitative measurement procedures.

© Clinical and Laboratory Standards Institute. All rights reserved.

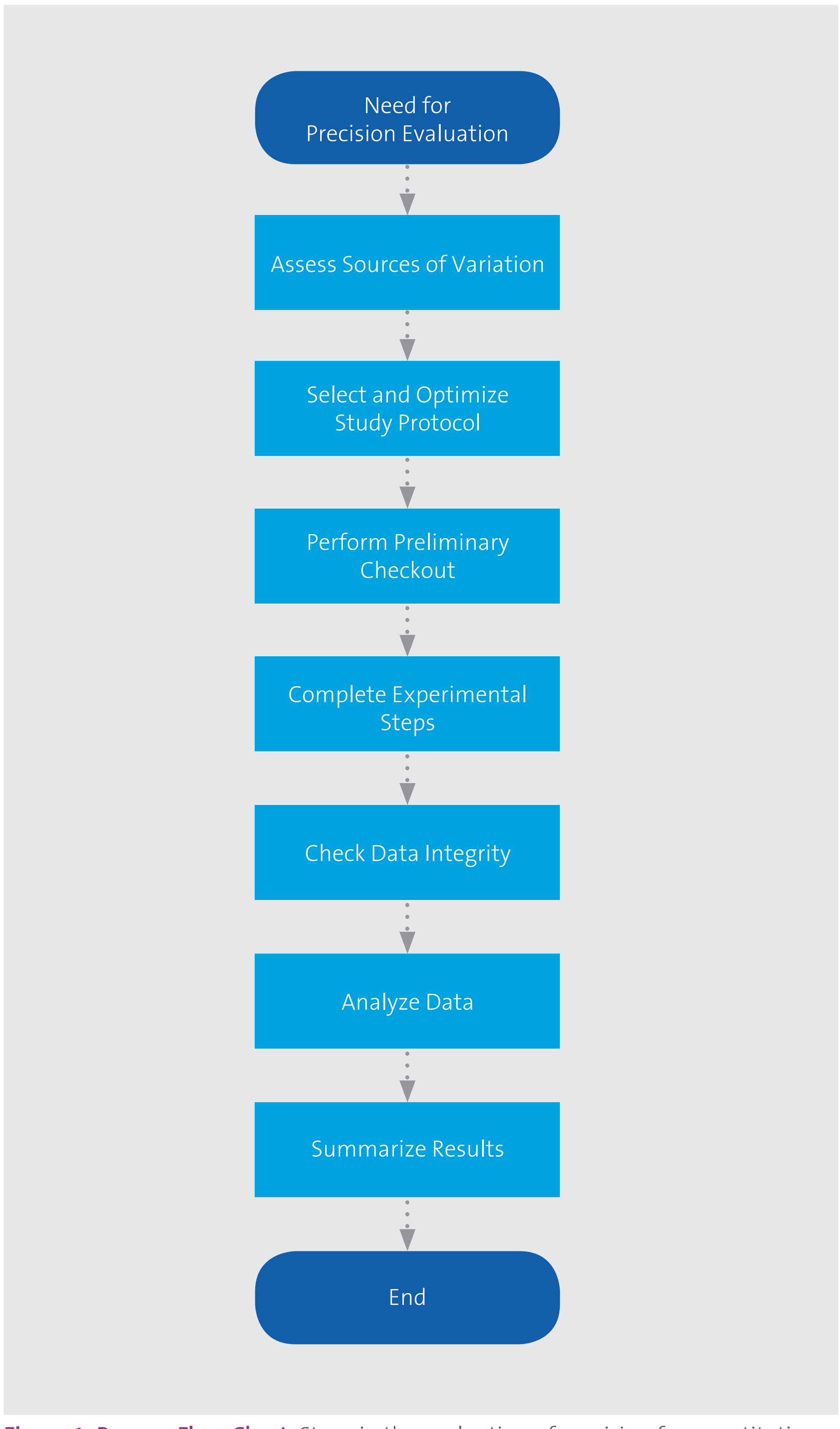


Figure 1. Process Flow Chart. Steps in the evaluation of precision for quantitative measurement procedures

© Clinical and Laboratory Standards Institute. All rights reserved.

1.5 Introduction to Basic Concepts

Both scientific and statistical expertise are essential to designing an appropriate experiment, or set of experiments, for establishing an assay's fundamental precision characteristics. Consider the task as collaboration between developer and statistician. As subject matter expert, the developer plays a major role. Even when a statistician assumes responsibility for technical details of the experimental design and analysis, the developer must possess at least a rudimentary understanding of relevant data analytical concepts in order to engage in constructive dialogue.

This section is addressed primarily to those with the relevant scientific expertise—assay developers, laboratory personnel, and others—who may not have a strong background in statistics or who wish to review the main statistical concepts relevant to analyzing the data. It provides an introduction to relevant terms and concepts, building on experience with routine QC in the clinical laboratory. The concepts featured here include:

- Variability and common measures of imprecision
- Sources of variation (sometimes called determinants or factors)
- Variance components and confounding
- Nested (ie, hierarchical) designs
- Balanced and unbalanced experimental designs and datasets
- ANOVA
- Statistical models and the symbolic notation commonly used for them
- Measures and indices for the uncertainty of precision estimates: confidence intervals (CIs), degrees of freedom (*DF*)

It is the **developer's responsibility to identify major sources of variation** associated with an assay.

This section also introduces the vocabulary of precision types adopted in this guideline—repeatability, within-laboratory precision, reproducibility—and comments on relationships to ISO terminology for precision and uncertainty. In addition, it sketches the two standardized experimental designs described more fully in Chapters 3 and 4.

The importance of the following items is stressed throughout the document:

- Identifying up front the major sources of variation associated with an assay.
- Customizing the design and size of the experiment(s), when necessary, to ensure that these sources are each exercised with sufficient thoroughness to yield suitably precise estimates of the precision types in question.

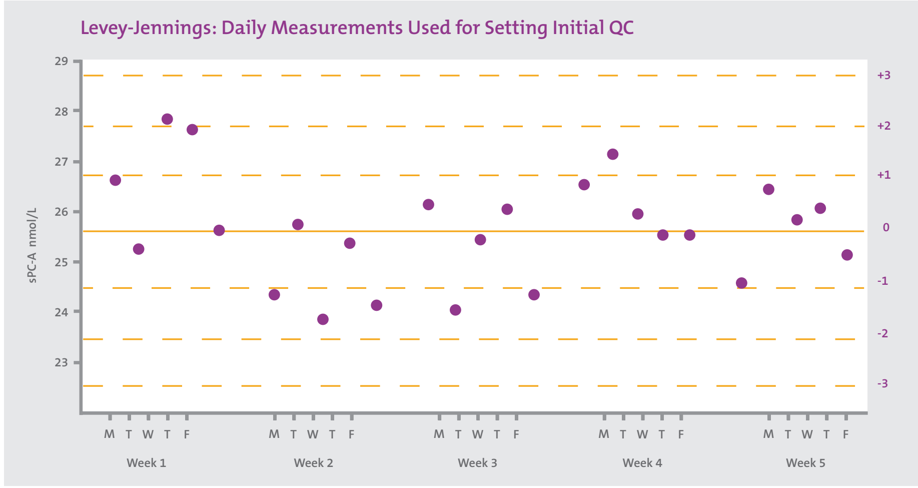
Chapter 2 provides guidance on optimizing the standardized experimental designs to reflect the major determinants of an assay's precision performance. For more complex situations, Appendix C provides an in-depth discussion of customizable, multifactor designs.

© Clinical and Laboratory Standards Institute. All rights reserved.

1.5.1 Variability and Sources of Variation

Quantitative clinical laboratory procedures characteristically display variability in results from repeated measurements. This should be familiar from internal QC and external proficiency testing (PT), as well as from simple experiments in which several replicates of a sample are tested in a single run.

Levey-Jennings charts,[17-19] widely adopted to monitor aspects of within-laboratory assay performance, display variability by plotting individual QC results on the vertical axis against time on the horizontal axis (see Figure 2). When the assay is performing as intended, the overall vertical dispersion of the points represents the assay's within-laboratory precision at the sample's concentration (or, more generally, its measurand level). That variability can be quantified by calculating mean and SD for a series of the QC results generated over a substantial period of time, eg, a month.



Abbreviations: M T W T F, Monday Tuesday Wednesday Thursday Friday; QC, quality control.

Figure 2. Levey-Jennings Plot. The graph depicts measurements generated to establish initial values (mean, SD) for a QC material. Concentration is represented on the left axis. A Levey-Jennings scale—representing deviations from the overall mean in SD units—is represented on the right axis. Time points are represented on the horizontal axis. (This graph is presented merely as an illustration of certain basic statistical concepts. Consult CLSI document C24[20] for guidance on internal QC.)

Other directly related measures of variability or dispersion are also in common use. The SD can be re-expressed as a variance by simply squaring it: *Var* = SD • SD. The variability can also be expressed in relative terms, when mean > 0, as a coefficient of variation expressed as a percentage: %CV = (SD / mean) • 100. Statistical output in the form of means and variances can always be converted into means and SDs, or means and %CVs, as needed.[a] (As a rule of thumb: fundamental calculations, such

NOTE:

As a rule of thumb: **fundamental calculations,** such as combining precision estimates, are best done in terms of variances, whereas **SDs** are more often appropriate for clinical interpretation, eg, in assessing the significance of a change or trend in serial measurements. And **%CVs** may be useful, for example, in graphically summarizing the concentration dependence of precision over a broad interval, as in Figures B3 and B4 in Appendix B.)

[a]In a simple experiment, the variance is $Var = \Sigma(x_i - x)^2 / (N-1)$. Note that this is the "unbiased" estimate of the variance and it is somewhat greater than the average of the squared deviations of the N results from their mean, because N – 1 rather than N figures in the denominator.

© Clinical and Laboratory Standards Institute. All rights reserved.

as combining precision estimates, are best done in terms of variances, whereas SDs are more often appropriate for clinical interpretation, eg, in assessing the significance of a change or trend in serial measurements. And %CVs may be useful, for example, in graphically summarizing the concentration dependence of precision over a broad interval, as in Figures B3 and B4 in Appendix B.)

For most assays, the variability exhibited in a single run is typically less than what is exhibited over the course of time in routine QC, while the variability exhibited in typical PT challenges is higher still. Intuitively, this is to be expected because there are more sources of variation contributing to within-laboratory precision than to within-run precision ("repeatability"), and still more contributing to interlaboratory precision than to within-laboratory precision.

By way of illustration, the dataset in Figure 2 exhibits apparent structure, in that the overall variability appears to arise from two distinct sources: a fairly homogeneous day-to-day variability—uniform in the sense that each of the five successive five-result clusters has essentially the same vertical dispersion—overlaid by random week-to-week variability. Designating the second source of variation as "weeks" accords with the pattern on display but has little explanatory force. More concretely, the shifts might reflect weekly recalibration or maintenance events, reagent changes, operator rotation, and so on, or more likely some combination of these and other determinants. The assay developer and laboratory personnel are in the best position to sort this out: they should consider all important sources of variability when designing the experiment(s) for evaluating precision.

Insofar as the week-to-week variability reflects a combination of two or more sources, those sources are confounded in the experimental design exemplified in Figure 2. In particular, the contributions of "days" and "runs" to the overall variability are confounded because the design involves just one run per day. In other words, the 5 × 5 experimental design corresponding to the set of QC results—five weeks, five days per week, one run per day—cannot resolve (ie, produce separate estimates for) these two sources of variation. Moreover, because the design involves just one replicate per run, it cannot yield an estimate of within-run imprecision (not that this is any shortcoming in the context of routine QC) even though the within-run determinants contribute to the overall variability.

For most clinical applications, **within-laboratory precision** (within-device precision) is typically of far greater relevance than repeatability (within-run precision).

1.5.2 Experimental Designs

This guideline describes two basic, standardized experiments appropriate for use with most quantitative assays in establishing or validating precision performance, and useful in other circumstances as well. These precision evaluation experiments are intended for estimating, in a controlled fashion, the variability observed in practice under three sets of conditions: 1) day-to-day within-laboratory, 2) interlaboratory, and 3) within-run conditions. In brief:

© Clinical and Laboratory Standards Institute. All rights reserved.

- The single-site study discussed in Chapter 3 and illustrated in Appendix A involves assaying multiple samples on 20 (not necessarily consecutive) days, with two runs per day, and two replicates per run: a 20×2×2 design. The statistical analysis treats the sources of variation associated with “days” and “runs” as factors (see Section 1.5.4); and it yields estimates for two precision types of interest: **repeatability** (within-run precision) and **within-laboratory precision** (within-device precision). For most clinical applications, the latter is typically of far greater relevance than repeatability.

  Suitable variants of the 20×2×2 design include designs with more than 20 days, designs with more than two replicates per run (for some or all days), and so on. For some devices, only one run per day may be needed. Or, with a particular measurement procedure, it may be possible to perform only one run per day. A design based on 20 days, with one run per day and three (or four) replicates per run—alternating, if possible, between morning and afternoon runs—would still allow for estimating the two precision types of interest. Compared to the 20×2×2 design, the 20×1×3 design would have effectively the same amount of data underlying the repeatability estimate. However, with only half the number of runs, the within-laboratory precision estimate would be expected to exhibit somewhat greater uncertainty; and a within-day between-run component of variance, if needed, could not be estimated, because day and run are confounded.

- The multisite study discussed in Chapter 4 and illustrated in Appendix B involves assaying multiple samples at three or more “sites” on five (not necessarily consecutive) days, with one run per day, alternating, if possible, between morning and afternoon runs, and five replicates per run: a 3×5×5 design. The statistical analysis treats “sites” and “days” (or “runs”) as factors, and it yields an estimate for a third precision type, namely **reproducibility.** This is particularly relevant for manual and partially automated assays, where operator-to-operator, instrument-to-instrument, and/or laboratory-to-laboratory variability may play a substantial role. The relatively brief standardized experiment is intended primarily to address doubts that might arise regarding the robustness of an assay’s precision performance across laboratories or instruments. Note that “day” and “run” are confounded in a multisite study involving just one run per day, which means that while their contributions are included in the reproducibility estimate, they cannot be estimated separately. This guideline does not require resolving or quantifying all of the sources of variation contributing to the precision estimates of interest. “Site” and “lot” would likewise be confounded if the three sites were each to use a different reagent lot—a generally undesirable study design, especially for new technologies, because it would then be impossible to determine from the results whether observed differences are lot related or site related.

  As noted in Chapter 4, a suitable alternative to the 3×5×5 design involves the same number of sites and days, but with two runs per day, and three replicates per run: a 3×5×2×3 design, which has effectively the same size for repeatability as the 3×5×5 design but requires a statistical analysis based on three factors rather than two. Other suitable variants of the 3×5×5 design involve simply increasing the number of sites and/or days. At minimum, an acceptable reproducibility study should involve at least three sites, at least five

“Day” and **“run”** are confounded in a multisite study involving just one run per day, which means that while their contributions are included in the reproducibility estimate, they cannot be estimated separately.

© Clinical and Laboratory Standards Institute. All rights reserved.

NOTE:

Periodic calibration provides an example; assays differ **both** in the frequency with which they require recalibration **and** in the degree to which calibration events represent a major or relatively minor source of variation. See additional discussion in Section 1.5.4.2.

days, and at least 20 *DF* for repeatability. (See Sections 1.5.3 and 1.5.4.2, respectively, for the concepts of repeatability and *DF*.)

Both the underlying sources of variation at work and their relative magnitudes may differ considerably from one assay to another, even for the same measurand. In particular, time—days in the 20 × 2 × 2 study, weeks in the 5 × 5 QC-based illustration—must be regarded as a surrogate for a number of more tangible determinants affecting variability. Standardizing the number of testing days cannot ensure that the same sources of variation are exercised, or exercised to the same degree, for all assays. Moreover, assays differ with respect to the relative influence (magnitude) of one source vs another. Periodic calibration provides an example; assays differ **both** in the frequency with which they require recalibration **and** in the degree to which calibration events represent a major or relatively minor source of variation. (See further discussion in Section 1.5.4.2.)

Accordingly, before settling on an experimental design or designs for evaluating an assay's precision performance, the developer should conduct an assessment—backed by knowledge and/or preliminary data for the assay in question and possibly for structurally similar assays as well—to identify those determinants representing the most important sources of variation. Chapter 2 provides guidance on how to apply this assessment in working out the details of a precision evaluation study. Appendix C addresses the design of more complex studies for assays where the two standardized experimental designs cannot do justice to the major sources of variation contributing to the precision types of interest.

1.5.3 Precision Types

The two standardized experiments in this guideline provide estimates for three clinically relevant precision types—repeatability, within-laboratory precision, and reproducibility—that are defined in terms of experimental conditions and/or the sources of variation associated with them:

- **Repeatability** (formerly called within-run precision) corresponds roughly to what could be estimated by calculating an SD directly from the results obtained by assaying several replicates for a given sample in a single run, in quick succession, with all experimental conditions held essentially constant. This accords with the notion of repeatability in the ISO classification scheme, because the measurements are collected over a short period of time, with all conditions—including operators, measuring system, location, etc.—kept the same as much as possible. Repeatability thus represents, in effect, the baseline "noise" of the assay: the lowest imprecision it is able to achieve in routine practice for a given concentration level. Some of the common sources of variation contributing to such "noise" are random variation in the pipetted volumes of sample and reagent; other sources are assay and/or system specific.

  In both of the standardized experiments, the repeatability estimate is derived, in effect, by "averaging" (in a certain way) across several independent within-run estimates, eg, the estimates expressed as

 © Clinical and Laboratory Standards Institute. All rights reserved.

variances obtained in 40 small runs (just two replicates each) in a single-site 20×2×2 study. For the evaluation of repeatability, this experimental design yields estimates of approximately the same uncertainty as a study based on a single run of N = 41 replicates (because 41 – 1 = 40 *DF*, as explained in Section 1.5.4.2), but is more assuredly representative, because there is no risk of its reflecting the idiosyncrasies of a particular run or day.

- **Within-laboratory precision** (aka "within-device" precision and formerly, in the first approved edition of EP05, called "total" precision) classifies as an "intermediate" precision type in the ISO framework because some conditions are held constant, but not all. In particular, the single-site 20×2×2 experiment described in this guideline holds site and instrument constant, while exercising both repeatability (ie, within-run) sources of variation and those additional sources of variation that impact the assay over a moderately extended period of time. Assuming a five-day work week, the study is expected to require about 26 to 28 days, or the better part of a month, to complete. Moreover, the number of days specified for the experiment happens to accord with common recommendations (ie, Westgard and CLSI document C24)[19,20] that initial values (SDs, as well as means) for QC materials be established via measurements on (at least) 20 distinct days.

  The within-laboratory precision type corresponds roughly, therefore, to what can be estimated from a series of internal QC results over a comparable time period. Indeed, simple measures of the overall variability in a dataset like the one illustrated in Figure 2 represent estimates of within-laboratory precision. Accordingly, SDs and %CVs determined in a single-site 20×2×2 study should be decent predictors of SDs and %CVs observed in routine QC, and vice versa, unless important sources of variation are represented in one context but not the other. Note, however, that it has been generally recommended (ie, Westgard and CLSI document C24)[19,20] that initial QC values eventually be replaced with statistics based on long-term cumulative data reflecting additional sources of variation represented only superficially or not at all in the initial assessment.

- **Reproducibility,** as explained in the Note on Terminology, encompasses sources of variation associated with multiple laboratories and/or multiple instruments, in addition to those sources affecting within-laboratory imprecision. This differs somewhat from the ISO characterization. A precision type classifies as a "reproducibility" precision type in the ISO framework only if it encompasses different locations **and** different operators **and** different measuring systems (instruments). On the other hand, the ISO framework presents no requirements for time, whereas it is implicit in the definition adopted here that a period of time is involved, encompassing at least five days.

NOTE on Measurement Uncertainty: An estimate of within-laboratory precision or even reproducibility at a given concentration level in a standardized EP05 imprecision study must be considered only a lower bound for the real-world imprecision relevant to routine clinical practice that includes the impact of certain long-term and very long-term sources of variation, such as reagent lots and calibrator lots, which will typically be underestimated in basic precision evaluation studies because

© Clinical and Laboratory Standards Institute. All rights reserved.

they are either excluded from the design or accorded, at most, token representation. Both for this reason and because additional sources must be taken into account for the estimation of so-called measurement uncertainty (additional sources that commonly would not be considered determinants of imprecision as such, as opposed to bias), an estimate of within-laboratory precision or even reproducibility must likewise be considered only a lower bound for the corresponding measurement uncertainty at that concentration level. Those additional sources of measurement uncertainty include systematic errors, some preexamination (preanalytical) errors not included in the standardized precision studies, patient sample-specific matrix effects (random interferences), etc. (see CLSI document EP29[21]). Measurement uncertainty as discussed in CLSI document EP29[21] characterizes an assay or individual results obtained by the assay. This must be distinguished from the notion of uncertainty—indexed by *DF* and measured by CIs (discussed in Section 1.5.4.2)—that characterizes the variability of precision estimates generated in a precision evaluation experiment or expected for a given experimental design.

1.5.4 Statistical Analysis

1.5.4.1 Making Sense of Statistical Models

Once the developer has assessed preliminary data to determine those sources of variation that need to be taken into account in evaluating the assay's precision, a suitable experiment(s) can be identified or tailored to reflect this assessment, and testing may get under way.

Statistical processing of the data, however, has to be based on ANOVA or some essentially equivalent methodology, which requires a symbolic formulation of how the data analysis is to treat composite sources of variation exercised in the study as contributing to the results obtained. As an illustration, consider a sample tested in a single-site 20 × 2 × 2 study. An individual, single-replicate result (Y_{ijk}) can be modeled as the mean (μ) for that sample plus random effects, ie, random deviations or "shifts" from the mean, where the deviations are regarded as attributable to two distinct factors, namely day-to-day sources of variation (D_i) and run-to-run within-day sources ($R_{j(i)}$), in addition to residual within-run sources ($\varepsilon_{k(ij)}$). The single-site 20 × 2 × 2 design calls for measuring each sample on n_{day} = 20 days, n_{run} = 2 runs per day, and n_{rep} = 2 replicates per run. Assuming that no results go missing, the final dataset will have altogether $N = n_{day} \times n_{run} \times n_{rep} = 80$ results per sample. The statistician models the 80-result dataset using, in effect, a set of N = 80 equations of the form:

$$Y_{ijk} = \mu + D_i + R_{j(i)} + \varepsilon_{k(ij)} \tag{1}$$

for $i = 1$ to n_{day}, $j = 1$ to n_{run}, and $k = 1$ to n_{rep}, where:

Y_{ijk} is the observed measurement for the *k*th replicate in the *j*th run on the *i*th day

μ is the mean, estimated as the average of the N = 80 measurements

© Clinical and Laboratory Standards Institute. All rights reserved.

D_i is the shift due to day-to-day sources of variation on the ith day

$R_{j(i)}$ is the shift due to run-to-run within-day sources in the jth run on the ith day

$\varepsilon_{k(ij)}$ is the shift due to within-run sources for the kth replicate in the jth run on the ith day

This describes a random-effects or components-of-variance model. The data analysis serves to fit the model represented by its formulation in equation (1), to the observed data, much as ordinary regression analysis fits a straight line model to method comparison data. That process yields statistics like slope and intercept, characterizing a relationship, whereas fitting a random-effects model yields estimates for variance components, which can be combined to produce estimates for recognizable precision types.

Nested models can also be called hierarchical models.

The nested (ie, hierarchical) structure of the model is shown graphically in Figure 3's top-most "comb" diagram, where replicates are nested within runs and runs are nested within days. This nesting is also indicated by parentheses in the subscripts of the model's symbolic formulation above (see Section 3.4).

All four of the experimental designs depicted in this figure are balanced. For the 20 × 2 × 2 protocol, this means that for a given sample, the design calls for assaying the same number of replicates in each run, and the same number of runs each day. It happens that the number of replicates per run

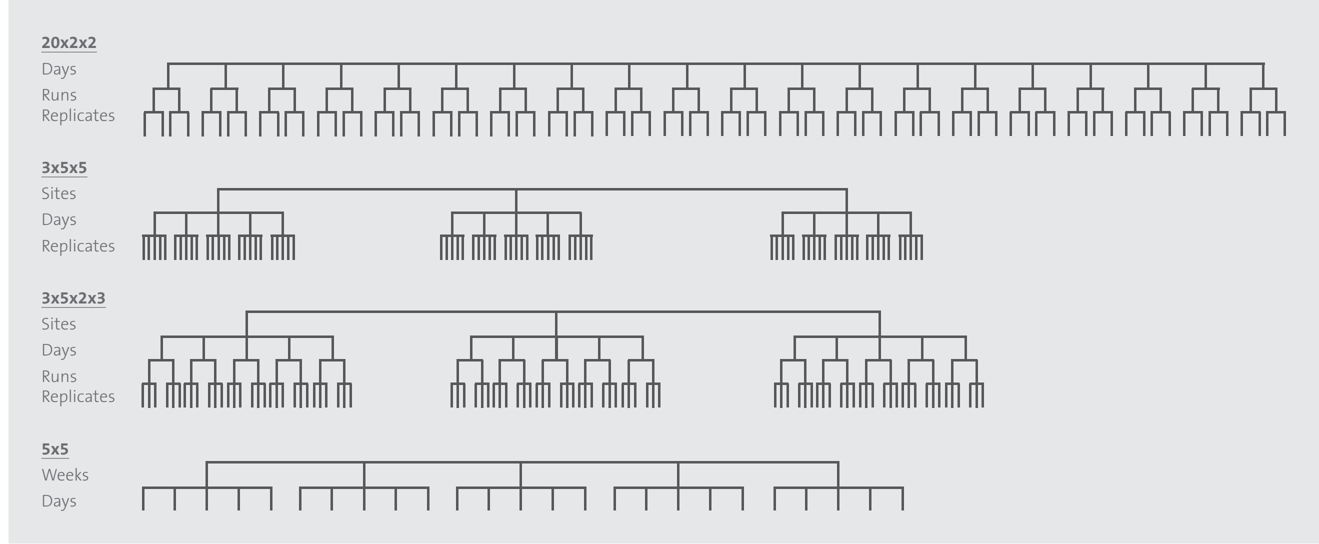


Figure 3. "Comb" Diagrams. These represent the standardized single-site 20 × 2 × 2 experimental design, the standardized multisite 3 × 5 × 5 and alternate 3 × 5 × 2 × 3 designs, and the QC-like 5 × 5 design used by way of illustration. All are nested (or hierarchical) structures, eg, for the 20 × 2 × 2, replicates are nested within runs, and runs within days. And all are balanced designs. The first and second designs each involve two factors—days and runs for the 20 × 2 × 2, sites and days for the 3 × 5 × 5—hence both designs are amenable to analysis by two-way nested ANOVA. The last two designs involve three factors and one factor, respectively, corresponding to three-way nested and one-way analyses.

© Clinical and Laboratory Standards Institute. All rights reserved.

is here the same as the number of runs per day, but that is not essential, as evidenced by the 3 × 5 × 2 × 3 design, which likewise constitutes a balanced design. Note that a study based on a balanced design may not yield a balanced dataset—for example, the dataset will be unbalanced if a result goes missing or is suppressed as a statistical outlier.

As depicted in Figure 4, random deviations due to variability at each stage in a single-site 20 × 2 × 2 study are modeled as contributions to the observed measurement result for an individual replicate. Runs and days are treated as random factors; what's left over ($\varepsilon_{k(ij)}$: the "residual error," or "error term") corresponds to within-run variability. For each sample, the classic statistical analysis—two-way (or "two-fold" or "two-level") nested ANOVA, or some equivalent methodology, for the standardized single-site 20 × 2 × 2 and multisite 3 × 5 × 5 studies—first dissects the total variability observed across all measurements obtained for a given sample, partitioning the variability into three variance components corresponding to the factors (here: "day" and "run") and the residual (ε). It then recombines these components to yield estimates for that sample of the precision types of interest.

For the single-site 20 × 2 × 2 study, the two-way analysis yields three variance component estimates for each sample:

- Variability attributable to within-run sources (ie, "residual error" or "measurement procedure noise")
- The additional variability attributable to within-day run-to-run sources
- The additional variability attributable to day-to-day sources

The within-run component serves as the repeatability estimate. The within-laboratory precision estimate is computed as a combination, specifically, as the total of all three variance components, hence the older designation of this precision type as "total precision." The term "total precision" is potentially misleading because, by definition, within-laboratory precision does not encompass interlaboratory sources of variation and, in practice, estimates of this precision type take little or no account of reagent and/or calibrator lot-to-lot and other potentially major sources of variation.

NOTE:

The distinction between **crossed and nested designs** is described in Appendix C.

Additional Notes on Statistical Modeling

- The reproducibility study adds a new term, *S*, for sites—ie, laboratories and/or instruments—while dropping the *R* term (unless more than one run is performed each day) and adjusting the subscripts accordingly. The reproducibility estimate is again computed as a combination of all three variance components. (See Section 4.4 for a symbolic specification of this model.)
- Models for designs featuring other sources of variation as factors, such as calibration events and reagent lots, can be constructed in a similar way by modifying the terms and subscripts. See Appendix C, which also deals with models involving crossed as well as nested designs. (For distinction between *crossed* and *nested* designs, see Section 1.3.2 and Appendix C.)

© Clinical and Laboratory Standards Institute. All rights reserved.

- Thinking of the illustrative QC dataset in Figure 2 as the output of a 5×5 experiment, the individual results can be modeled as the mean plus a random shift "due to weeks" (the W_i term) and another representing residual, within-week variability (the so-called error term, $\varepsilon_{j(i)}$):

  $$Y_{ij} = \mu + W_i + \varepsilon_{j(i)} \quad (2)$$

  for $i = 1$ to n_{week}, $j = 1$ to n_{day}, where $n_{week} = n_{day} = 5$.

  Because there is only one factor (not counting the residual "what's left over" source), the natural method for processing this data is one-way ANOVA. ANOVA yields estimates for two variance components, which can be combined to yield an estimate of the precision type of principal interest in the QC context.

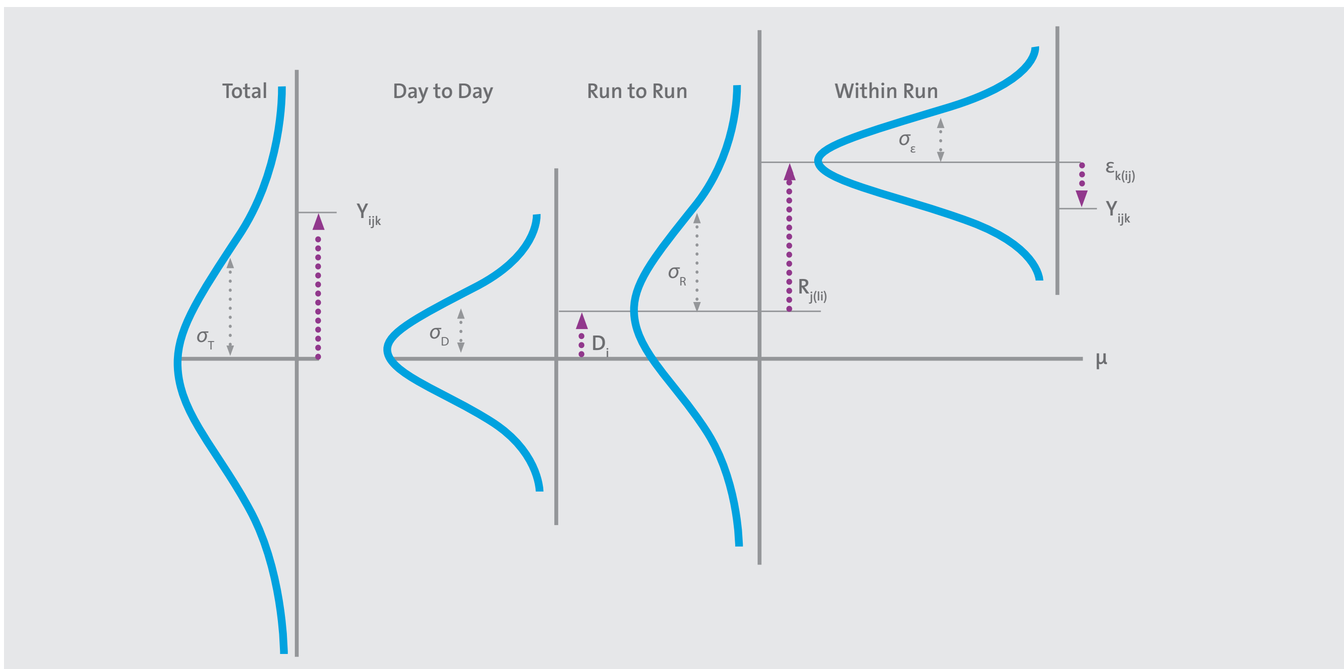


Figure 4. Variance Decomposition. An Individual Result Y_{ijk} (ie, the result for a single replicate) in a 20×2×2 study, modeled as the vector sum of deviations—$D_i + R_{j(i)} + \varepsilon_{k(ij)}$—from the sample's mean, μ, due to the combined effect of three composite sources of random variation—namely, day-to-day (D), run-to-run within-day (R), and within-run or residual (ε) variance components—with distributions characterized by SDs of σ_D, σ_R, and σ_ε, respectively. The heavy, single-headed arrows represent the magnitude and direction of the overall (total) and three source-specific deviations (shifts). The thin, double-headed arrows represent the magnitudes of the respective SDs.

1.5.4.2 Confidence Intervals and Degrees of Freedom

The statistical analysis yields not only estimates—(single-valued) point estimates—for both the variance components and combinations thereof corresponding to the precision types of interest, but also CIs bracketing these estimates. A CI may be considered a (two-valued) interval estimate for the magnitude of the variance component or precision type in question, its width indicating how narrowly the experiment has succeeded in pin-pointing that magnitude. CIs serve to characterize the imprecision or uncertainty of the estimates as far as experimental "sampling variation" is concerned. Thus CIs can be controlled to some extent by making suitable

© Clinical and Laboratory Standards Institute. All rights reserved.

adjustments to the experiment's size: an experiment that is larger in relevant dimensions can be expected to reduce the impact of sampling variation, thereby yielding more precise estimates as exemplified by correspondingly narrower CIs.

For a given confidence level (CL)—conventionally 95%— the CI for a precision estimate is determined by (a) the point estimate itself and (b) what statisticians call *DF* (its degrees of freedom), a number that may be thought of as indexing the amount of data underlying the estimate: the effective size of the experiment as it relates to the estimate in question.[b] In the simplest case, *DF* = N – 1. (Recall that the "N – 1" also figures in the formulae for variance and SD.) This holds for a dataset with N values and no apparent structure, a dataset that can be regarded therefore as arising from a homogeneous process, a single source of random variation. Here the interpretation is straightforward. The CI's width is a measure of the estimate's uncertainty. A larger experiment (increased N-size) can be expected to yield a more precise, more informative precision estimate, ie, an estimate with a narrower CI, hence less uncertainty.

Both N-size and *DF* thus serve as natural indices of the imprecision or uncertainty of the precision estimates.

For precision studies, the *DF* concept becomes distinctive, and important in its own right, when indices of experimental size must be assigned to more than one source of variation or precision type. For example, when the apparent structure for the QC data in Figure 2 is taken into account, there are two distinct sources of variation: week-to-week and within-week. Also of interest is their "sum," which corresponds to what the direct calculation of an SD for the dataset as a whole would be aiming to assess. The dataset is inherently multidimensional, requiring more than one number to characterize its size. In this illustration, two numbers are required, one for each of the variance components or sources of variation involved.

- For week-to-week variability—the source of variation that would be treated as the factor in a one-way ANOVA—there are just 5 – 1 = 4 *DF* in the five-week study depicted in the graph, in accord with the *DF* = N – 1 relationship. The *DF* for a variance component or its estimate indicate how extensively that source of variation is exercised in the study.
- For the residual within-week variability, there are 20 *DF*, reckoned as follows. The impact of "weeks" can be factored out by regarding the study as a succession of five smaller (single-week) experiments, each with five results and no apparent structure, and each yielding an estimate of within-week variability with 4 *DF*. Combining across the five independent estimates expressed as variances involves computing a certain weighted average to produce a more definitive estimate of within-week variability associated, in aggregate, with 5 • 4 = 20 *DF*. (Combining estimates in this manner is explained in Section 3.7.1.)

[b] Specifically, for an estimate expressed as an SD, the upper and lower confidence limits are computed as the SD times the square root of (DF / X), where X is the chi-square value tabulated for *DF* and the relevant quantile, namely 0.025 or 0.975 for a central 95% CI.

© Clinical and Laboratory Standards Institute. All rights reserved.

Accordingly, making use of the N = *DF* + 1 relationship, the estimate of within-week variability (associated with 20 *DF*) from the five-week study can be expected to have essentially the same imprecision as an estimate of the same magnitude obtained in a study generating N = 21 results in a single week, because the CIs would be identical. (Of course, the estimate obtained from the longer study would be more assuredly representative. CIs reflect only sampling variation.)

NOTE:

Figure 5 offers a quantitative example of how the **direct and components-of-variance approaches** compare in this respect.)

As for overall variability in the QC example depicted in Figure 2, ANOVA calculates the "total" SD as a combination of the week-to-week and within-week variance component estimates, rather than by computing an SD directly from the 25 results as if there were only one, homogeneous source of variation. The direct approach can usually be expected to yield nearly the same precision estimate, but it would attribute, with unwarranted optimism, 25 – 1 = 24 *DF* to the estimate. The components-of-variance approach differs primarily in being more conservative and realistic in its assessment of the uncertainty associated with the precision estimate. Compared to the direct approach, it ascribes fewer *DF* to the estimate, and this translates into a wider CI, signifying greater uncertainty. (Figure 5, to be discussed shortly, offers a quantitative example of how the direct and components-of-variance approaches compare in this respect.)

NOTE:

See Table 2 in Section 3.6.2 and Table A2 in Appendix A for examples of ANOVA tables.

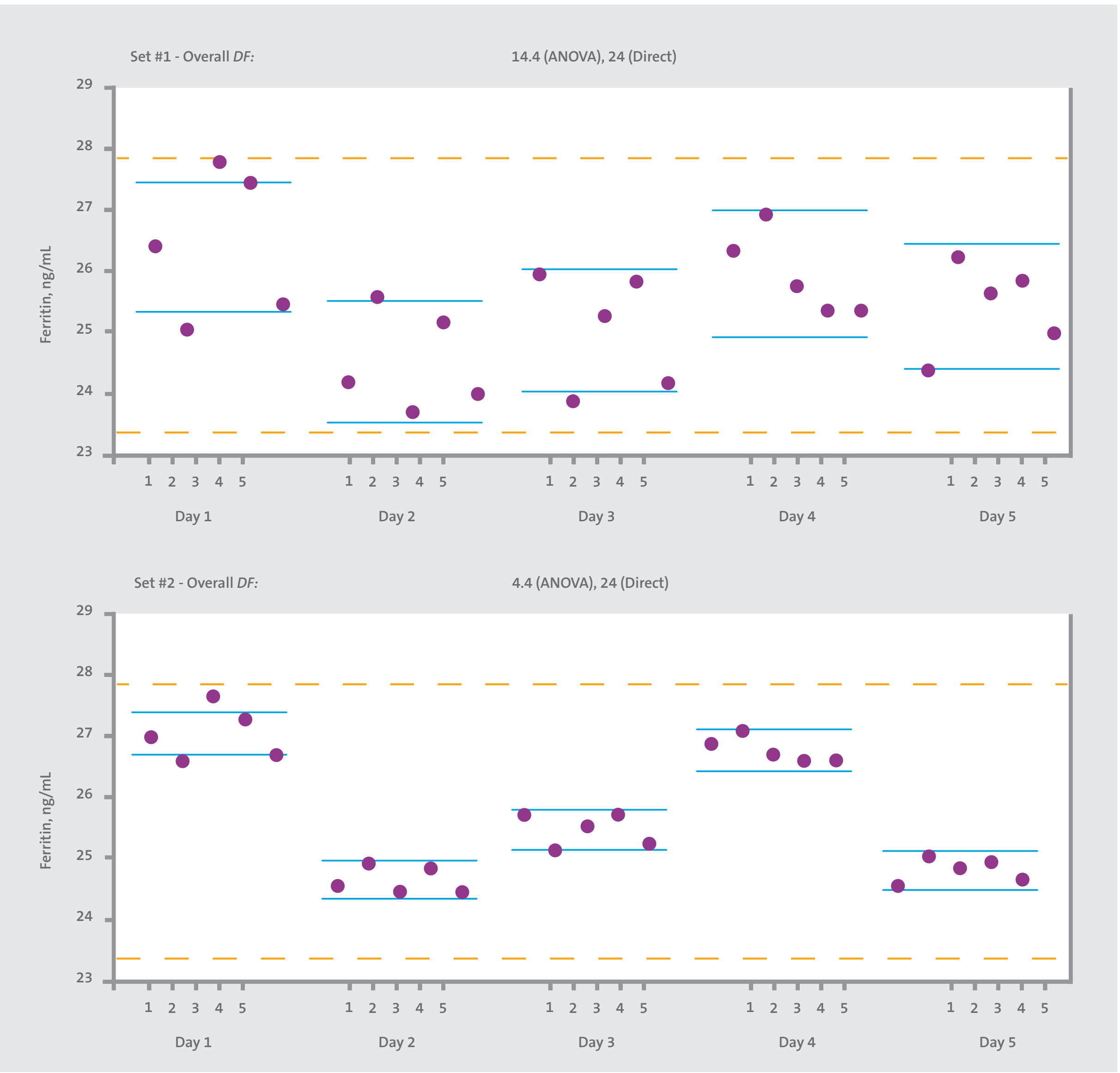
Because the precision estimate represents a combination of variance components, the *DF* associated with it must be approximated (eg, by Satterthwaite's method)[22] in the course of the analysis, unlike the *DF* for the two components (the week-to-week and within-week sources of variation), which can be read off in advance from the statistical model associated with the experimental design. (See Table 2 in Section 3.6.2 and Table A2 in Appendix A for examples of ANOVA tables.)

For the estimate of a precision type (such as within-laboratory precision) representing a combination of variance components, a contributing source of variation will affect the estimate's *DF*. As such, it will also affect its imprecision or uncertainty as measured by the width of the estimate's CI, depending on:

- How frequently the source is exercised in the study
- How important it is, ie, its magnitude relative to the other sources of variation figuring in the experiment.

Figure 5 illustrates the principles involved.

© Clinical and Laboratory Standards Institute. All rights reserved.



Abbreviations: ANOVA, analysis of variance; *DF*, degrees of freedom.

Figure 5. Contrasting Datasets. These exhibit essentially the same overall variability but differ markedly in the relative contributions from within-day and day-to-day sources of variation. The frames superimposed on the data points represent the ± 2 SD dispersion for within-day variability (solid lines) and overall variability (dashed lines), as estimated by one-way ANOVA.

In certain respects, the two datasets in Figure 5—consisting of measurements generated by assays from two different developers—are well matched. They have the same mean and, when calculated directly, the same overall SD. They differ just slightly in overall SD when that is estimated by one-way ANOVA, the most elementary components-of-variance analysis. Moreover, for each of the two datasets, the direct and ANOVA approaches yield nearly the same estimates for overall SD. (This is typical, but the estimates must always be expected to differ somewhat, due to differences in the calculations involved.)

© Clinical and Laboratory Standards Institute. All rights reserved.

Mean and SD do not tell the whole story, however: there is an important difference between the two datasets. For Set #2, in contrast to Set #1, variability "due to days" is visibly dominant, accounting for most of the variability on display; and this apparent structure cannot readily be dismissed as spurious, as it might be for Set #1. The structural difference is reflected in the components-of-variance analysis, but not in the direct analysis. In general, the most striking difference between the direct and ANOVA approaches lies not in the overall precision estimates, but rather in the *DF* attributed to those estimates, and thus in the resulting CIs as well. Compared to the direct approach, one-way ANOVA yields a more conservative—less optimistic—assessment of the quality of the overall precision estimate for Set #2, that is, a far lower *DF* and consequently a wider CI. This accords better with the sense that much greater uncertainty attaches to the overall SD for the second dataset compared to the first. Indeed, for Set #2, the *DF* estimated by ANOVA for overall precision is just marginally higher than the *DF* associated with the day-to-day component of variance (4.4 vs 4), implying that the effective size of this dataset is comparable to that which might arise from a study generating just one result per day over five or six days.

NOTE on Figure 5: Both datasets could materialize from a five-day study, based on one run per day and several measurements per run, such as a standardized reproducibility study (see Chapter 4) or a precision verification study from CLSI document EP15.[1] Both datasets have 20 *DF* and 4 *DF*, respectively, for the within-day and day-to-day sources of variation. When analyzed directly, ignoring structure, both have an (overall) SD of 1.07 ng/mL with 24 *DF* and a 95% CI of 0.84 to 1.49 ng/mL. For the first, relatively homogeneous dataset, ANOVA yields an overall ("total" or "within-laboratory") SD of 1.10 ng/mL with 14.4 *DF* and a 95% CI of 0.81 to 1.72 ng/mL; whereas for Set #2, it yields an overall SD of 1.17 ng/mL with just 4.4 *DF* and a broader 95% CI of 0.71 to 3.13 ng/mL. Note, incidentally, that fractional *DF* values, like 4.4, as opposed to integer values, like 4, are typical of methods, like Satterthwaite's,[22] for approximating the *DF* of "linear combinations" (weighted sums) of variance components. (For more on Satterthwaite's method, see Sections 3.6.3 and 4.6.3.)

This is of great importance when designing studies for estimating precision types, like within-laboratory precision and reproducibility, which represent linear combinations of variance components, ie, multiple sources of variation combined. To the extent that a source represents a major determinant of the combined variability being targeted, it should be accorded more than token representation, because the quality of the precision estimate will depend on the frequency (reflected in the *DF* for that source implied by the experimental design) with which the source is exercised in the study. Accordingly, the manufacturer or developer should have a good sense of the source's magnitude, relative to that of other sources of variation, in advance of planning the study.

© Clinical and Laboratory Standards Institute. All rights reserved.

EXAMPLE: As a practical example, consider how calibration events can affect single-site 20 × 2 × 2 experiments. Assume here that the experiments are completed in 26 to 28 calendar days, and that the assays are calibrated on "day one" and thereafter (as recommended in this guideline) according to the manufacturer's stated requirements for periodic recalibration.

- For an assay recalibrated every day, the overall variability ("within-laboratory precision") assessed in the 20 × 2 × 2 study will substantially reflect the contribution, great or small, from calibration events, because there are altogether N = 20 such events, corresponding to 19 *DF* for that source of variation.
- For an assay recalibrated on a monthly basis or still less frequently, the impact of recalibration will not be exercised at all in the 20-day study, and yet this may be a substantial source of clinically relevant long-term variability. N = 1 corresponds to zero *DF*, which makes sense intuitively as no variability due to multiple calibration events is represented in the study's experimental design or dataset. (In reporting the study, it should be made clear that the within-laboratory precision estimate does not incorporate variability due to calibration events.)
- For an assay recalibrated every two weeks, only N = 2 such events will figure in a typical 20 × 2 × 2 study, which translates into just 1 *DF* for this source of variation. If it is a relatively minor source, this will have no adverse effect on the study's estimates for the precision type of principal interest. (Remember that precise estimates, ie, estimates associated with narrow CIs, for the variance components [other than the "within-run" or "residual" component corresponding to repeatability] are not among the study's goals.) Whereas if calibration events represent a major source of variation, this will compromise the study's estimate of within-laboratory precision, resulting more often than not in a potentially serious underestimate of that precision type, but sometimes in a considerable overestimate. In short, for this scenario involving infrequent calibration events, the greater the magnitude of the variability due to that source, the greater the uncertainty associated with the estimate obtained for within-laboratory precision.

  Of course, the uncertainty in this worst case (*DF* = 1) scenario could be reduced by performing recalibration more frequently than the assay's requirements for periodic recalibration would dictate, though the desirability of this expedient might well be offset by other considerations. In particular, recalibrating on (for example) a daily basis could mask the drift characteristic of the recommended period between calibration events. In principle, the dataset could be re-analyzed, extracting an estimate of that drift, treating it as an additional variance component, and incorporating it into the estimate for within-laboratory precision. But this would involve a complex and far from transparent calculation, hence it is not recommended. Calibration events should figure in the experimental design at the very beginning of the study and thereafter only at the frequency dictated by the assay's stated calibration period, ie, at the rate intended for recalibration in the field by clinical laboratories.

© Clinical and Laboratory Standards Institute. All rights reserved.

Subjecting the samples to several 20 × 2 × 2 studies—each representing a different combination of reagent lot and instrument, for example, as often happens in the course of validating a new assay—and combining the statistics across these studies in a suitable way will have a beneficial impact on the situation, substantially increasing the *DF* due to calibration events. Within-laboratory precision estimates generated in this manner will reflect variability due to calibration events more accurately than the statistics from any one of the studies involved, in addition to being more representative across reagent lots and instruments.

Note that all three scenarios are associated with the same ANOVA model, encapsulated in Section 1.5.4.1, equation (1). Calibration events figure not as a distinct factor in this statistical model, but rather as one of presumably many sources of variation contributing to the variance component labeled "days." Accordingly, not all facets of the experimental design are settled by the choice of statistical model. Relevant considerations are addressed more fully in the following section.

© Clinical and Laboratory Standards Institute. All rights reserved.

This page is intentionally left blank.

© Clinical and Laboratory Standards Institute. All rights reserved.

Chapter 2

Selecting and Optimizing an Appropriate Precision Evaluation Protocol

This chapter includes:

- Guidance on designing a study for evaluating the precision performance characteristics of a measurement procedure
- Factors for consideration when selecting and optimizing an appropriate precision evaluation protocol
- Sources of variation for common measurement procedures
- Introduction to two standardized precision evaluation study designs

2 Selecting and Optimizing an Appropriate Precision Evaluation Protocol

This chapter provides guidance to manufacturers and developers on designing a study for evaluating the precision performance characteristics of a measurement device. (The study may well consist of multiple experiments, each applied to samples representing a spectrum of measurand levels.) In deciding on the study's architecture and details, dialogue with the appropriate regulatory agency may be advisable, when relevant. In any case, the considerations supporting the decision should be documented, and they should reflect the sources of variation affecting the assay, and their relative magnitudes and interrelationships. For a given assay, the most appropriate study—its design (statistical model), optimization (implementation details), and reporting—will depend on the assay's character and also, to some extent, on its intended use.

IMPORTANT NOTE:

The study described in this section may consist of **multiple experiments,** each applied to samples representing a spectrum of measurand levels.

A good starting point for the study design process is consideration of the standardized 20 × 2 × 2 framework. (This was introduced in Section 1.5 and will be discussed at length in Chapter 3.) This nested two-factor random-effects model, featuring **days** and **runs** (runs nested within days), has served well, since the first approved edition of EP05, as the basis for assessing within-laboratory precision (and repeatability, as well) for the majority of quantitative assays used in the clinical laboratory. It can be adapted to the needs of a particular assay in four different ways to ensure that sources of variation important to the assay are appropriately addressed in the precision study:

NOTE:

In deciding on the **study's architecture** and details, dialogue with the appropriate regulatory agency may be advisable, when relevant.

- The factors can be changed, eg, from **days** and **runs**, to **sites** and **days**. Structurally, this is the fundamental difference between the single-site 20 × 2 × 2 and multisite 3 × 5 × 5 designs, which, size apart, exemplify the same two-factor statistical model. (The 3 × 5 × 5 design and variants thereof are discussed in Chapter 4.)
- One or more additional factors, representing, for example, operators, reagent lots, and/or calibration events, can be introduced into the model. See Appendix C for an in-depth discussion of designs involving three or more factors.
- Benchwork-related details of the 20 × 2 × 2 protocol can be modified to ensure that a particular source of variation (eg, operator variability) is incorporated into the within-laboratory precision estimate via its impact on one factor or the other. Note that this does not amount to treating the source in question as a separate factor in the experimental design or analysis.
- The core 20 × 2 × 2 experiment(s) can be supplemented with one or more experiments dedicated to some additional source(s) of variation. This is the primary role intended for the 3 × 5 × 5 study, which is designed to address site-to-site variability.

© Clinical and Laboratory Standards Institute. All rights reserved.

For addressing a given source of variation, there may be several viable options. Consider operator-to-operator variability. The simplest and most natural approach, corresponding to the third option listed above, would be to vary the operator from run to run or from day to day. Either way, the variability would be incorporated into the estimates of within-laboratory precision, as is vital when evaluating the precision of a manual or semiautomated assay. In actual practice, the protocol might call for randomly assigning responsibility to one of several available technicians for each run or for each day's benchwork. In summarizing the study, the number of operators should be reported and also the mode of randomizing them across runs or days. This should suffice to document how well the precision estimates obtained in the study reflect operator-to-operator variability. (Alternative approaches, corresponding to the second and fourth options, might also be considered. At the likely cost of increased benchwork and increased data processing complexity, the 20 × 2 × 2 experiment could be augmented by introducing "operator" as a third factor, or a separate experiment dedicated to addressing operator-to-operator variability could be made part of the precision study.) If the assay under investigation is fully automated and operator-to-operator variability can be regarded as negligible, this may justify both ignoring operator-to-operator variability in the study design and dropping mention of the number of operators, etc., in the report.

As another example, consider reagent lot-to-lot variability. When an assay is first introduced and ready for subjecting to validation studies, there are rarely more than a very few reagent lots available for testing and these can hardly be considered a "random sample." At this stage, therefore, there is no way to obtain for the assay solid estimates of the component of variance associated with lot-to-lot variability. Nevertheless, it is possible to ensure that within-run precision estimates generated for the assay are at least somewhat representative across reagent lots, even though these estimates do not include variability from this potentially major source of variation. Approaches corresponding to the second and fourth options listed above and involving different reagent lots might be considered: repeatability, between-run, between-day, and between-lot components should be evaluated. It is important to report the experimental design and data analysis in a way that makes it clear what is actually included in the within-laboratory precision estimates.

As a third example, consider instrument-to-instrument variability. To account for this source of variation, "instrument" could be introduced into the basic 20 × 2 × 2 model as an additional factor. Alternatively, a supplementary multisite 3 × 5 × 5 study could be performed, with "site" interpreted as "instrument." Either of these approaches would produce an estimate of the instrument-related variance component.

As illustrated for operator-to-operator, reagent lot-to-lot, and instrument-to-instrument variability in the three preceding paragraphs, the familiar

Alternative approaches, corresponding to the second and fourth options, might also be considered. At the likely cost of increased benchwork and increased data processing complexity, the 20 × 2 × 2 experiment could be augmented by introducing "operator" as a third factor, or a separate experiment dedicated to addressing operator-to-operator variability could be made part of the precision study.

© Clinical and Laboratory Standards Institute. All rights reserved.

two-factor experimental framework, suitably optimized, can be expected to play a pivotal role in precision evaluation studies for a wide variety of assays.

2.1 The Need for Multiple Study Options

> **NOTE:**
> It is understood that there may be situations where **different protocols and/or different statistical techniques** may be appropriate. Not described in this document, for example, are robust and nonparametric statistics, although Appendix B, Section B4 illustrates one application for such techniques in screening for potential outliers in a precision study.

The extremely wide range in design, implementation, and operation of IVD devices makes it impractical to recommend a single experimental protocol for the precision evaluation of all quantitative clinical laboratory measurement procedures. Instead, this document describes multiple protocol options designed to account for different sources of variation that may be relevant to the evaluation of precision for a given measurement procedure. The manufacturer or developer may choose to use one or more of these options, or devise an alternative protocol, depending on the character of the measurement procedure's design, and sometimes on its intended use as well, which could render certain sources of variation more important or less important to address.

All of the statistical analyses described or referenced in this document can be performed with well-established methods for components-of-variance analysis—a family of techniques encompassing ANOVA, restricted maximum likelihood (REML), minimum norm quadratic unbiased estimation (MINQUE), minimum variance quadratic unbiased estimation (MIVQUE), and other approaches.[23] Because different analytical techniques and different software implementations may produce somewhat different numerical estimates of the variance components, their sums, and their confidence limits, reports should include relevant information on the statistical technique(s) and software implementation(s) used for the analysis.

It is understood that there may be situations where different protocols and/or different statistical techniques may be appropriate. Not described in this document, for example, are robust and nonparametric statistics,[24] although Appendix B, Section B4 illustrates one application for such techniques in screening for potential outliers in a precision study.

2.2 Study Design Selection Considerations

Selecting and optimizing an appropriate precision evaluation protocol for any particular measurement procedure depends on:

- Deciding which precision types are to be evaluated
- Knowing which sources of variation might have a major impact on those precision types

© Clinical and Laboratory Standards Institute. All rights reserved.

These two aspects must be considered jointly. For example, while many sources of variation, including reagent and calibrator lot-to-lot variability, may contribute substantially to the overall long-term precision of an assay, some of these sources may not be relevant to typical within-laboratory precision estimates, such as those generated via the standardized 20 × 2 × 2 design, discussed in Chapter 3, which is reflective of the assay's use over a substantial but more limited timeframe consisting of approximately one month.

2.2.1 Precision Types to Be Evaluated

The primary types of precision estimated for a clinical laboratory measurement procedure are repeatability (formerly known as within-run precision) and within-laboratory precision (formerly known as "total" precision). As described in Section 1.5, **repeatability** reflects the variability among replicate measurements of a sample under experimental conditions held as constant as possible—effectively the baseline "noise" of a measurement procedure. **Within-laboratory precision** (aka, within-device precision—in ISO terms an "intermediate" precision type), builds upon repeatability by incorporating run-to-run and day-to-day sources of variation on a single instrument. (Depending on the experimental framework adopted, and how it is implemented, many specific sources of variation may be incorporated in this precision type, eg, operator-to-operator variability for manual or semiautomated procedures, calibration cycle-to-cycle variability for procedures that require frequent calibration.) As such, repeatability and within-laboratory precision are evaluated in a single-site study, such as the standardized 20 × 2 × 2 design discussed in Chapter 3.

Depending on the particular measurement procedure and its clinical applications, it may also be appropriate to provide an estimate of reproducibility. As defined in this document, **reproducibility** incorporates laboratory-to-laboratory variability and/or instrument-to-instrument variability on top of the sources contributing to within-laboratory precision (see Section 1.5). As such, reproducibility is addressed in a multisite (multilaboratory or multi-instrument) study, such as the standardized 3 × 5 × 5 design discussed in Chapter 4.

More comprehensive statistical models can provide estimates of specific sources of variation (eg, instrument-to-instrument, reagent lot-to-lot) as well as estimates for additional variance components and intermediate precision types. These may be of great interest to manufacturers for internal product performance characterization and for prioritizing improvements.

REMINDER:

Repeatability was formerly known as **within-run precision. Within-laboratory precision** was formerly known as **total precision.**

IMPORTANT NOTE:

Depending on the **experimental framework** adopted, and how it is implemented, many specific sources of variation may be incorporated in this precision type, eg, operator-to-operator variability for manual or semiautomated procedures, calibration cycle-to-cycle variability for procedures that require frequent calibration.

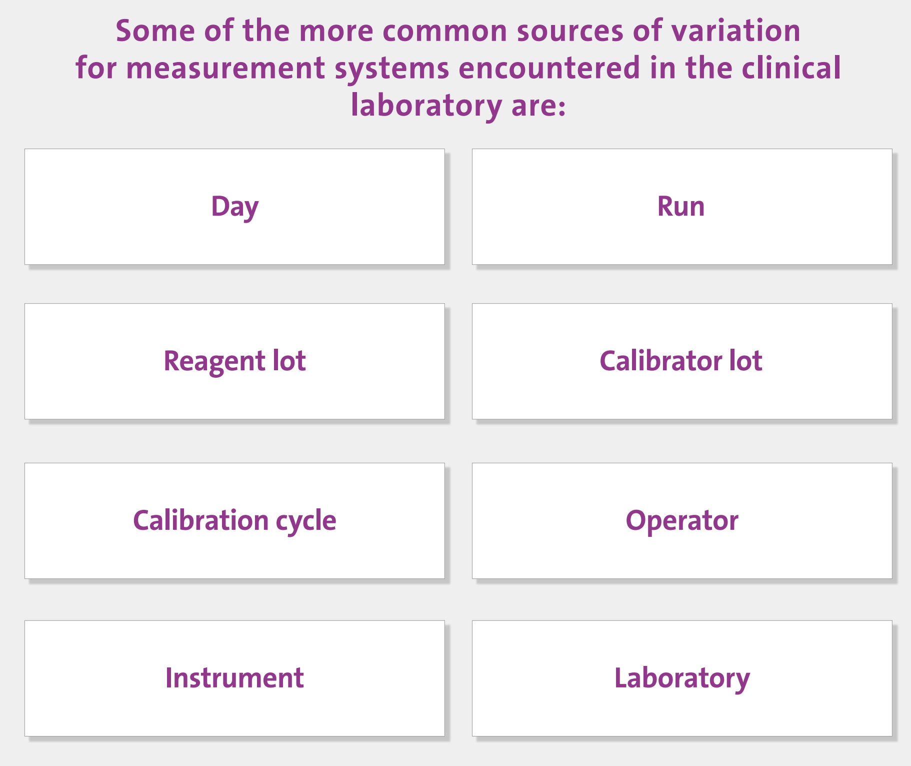
REMINDER:

Reproducibility incorporates laboratory-to-laboratory variability and/or instrument-to-instrument variability on top of sources contributing to within-laboratory precision.

© Clinical and Laboratory Standards Institute. All rights reserved.

2.2.2 Sources Contributing to Measurement Procedure Variability

Many different sources can contribute to the variability of results from a measurement procedure. The particular sources and the extent and clinical relevance of their impact will vary from one assay or measurement device to another, possibly even for different applications of the same measurement procedure, eg, cancer marker measurements used for diagnostic support vs for monitoring.



Not all of these sources may be relevant for a given measurement procedure; on the other hand, one must be wary of dismissing a source as irrelevant. For example, it might be thought that consideration of operator-to-operator variability is not relevant for a so-called "fully automated" chemistry procedure, and yet such variability might actually exist and be substantial due to operator influence on specimen handling or other relevant preexamination (preanalytical) elements of the measurement procedure.

Some determinants of variability such as sample preparation, test material stability, carryover, and drift (see CLSI document EP10[25]) are not estimated separately but are understood to influence within-laboratory precision estimates.

Some factors, such as "day" and "run" in the 20 × 2 × 2 design, may be surrogates for a variety of determinants. That is, many sources of variation may be subsumed under, or collectively contribute to the magnitude of, a given factor or variance component figuring in a statistical model. Thus,

© Clinical and Laboratory Standards Institute. All rights reserved.

site-to-site variability in a multisite precision evaluation may be more properly due to calibrator lot-to-lot variability (for calibrated assays) or instrument-to-instrument variability (mostly for noncalibrated assays), unless there are specific contributions from local environmental conditions (eg, altitude, humidity). Likewise, some sources of variation may be included in others, so that assessment of the latter thereby accounts for the former as well. For example, the 20 × 2 × 2 and 3 × 5 × 5 designs both yield estimates of within-laboratory precision, even though the 20 × 2 × 2 includes factors for both days and runs, while the 3 × 5 × 5 treats only one of these (days) as a factor. But day-to-day variability encompasses virtually all of the sources of variation associated with runs, except possibly for time-of-day variability, which could be introduced into the 3 × 5 × 5, without increasing the number of factors involved, by alternating between morning and afternoon runs or otherwise "randomizing" the time of day for the daily runs.

Product developers can investigate the relevance and magnitude of sources of variation during assay development through a variety of approaches, such as:

- Literature reports
- Competitive product information
- Feedback from consultants and expert users
- Experience with similar assays
- Risk analysis techniques, such as product and process failure modes and effects analysis
- Research and development, production, or QC data for the assay under investigation
- Precision studies for the assay performed before the precision study now being planned

The purpose of this investigation is to determine which sources of variation can be expected to have a major impact on the precision of the measurement procedure being evaluated, thus indicating which sources to address in the precision study's design, either as separate factors in the statistical model or, more commonly, as elements contributing to (or incorporated in) the variance components assessed in a standard model, eg, as elements subsumed under the "day," "run," or residual terms in a 20 × 2 × 2 design. As an initial guide, Table 1 lists the sources of variation typically associated with repeatability and/or within-laboratory precision for common types of measurement procedures, and makes note of typical inclusion relationships. Note that when an estimate of the assay's reproducibility is desired, "site" (laboratory or instrument) must be one of the factors in the statistical model, as illustrated by the multisite 3 × 5 × 5 design discussed in Chapter 4.

© Clinical and Laboratory Standards Institute. All rights reserved.

Table 1. Sources of Variation for Common Types of Measurement Procedures

Example	Device Type	Example Methods	Potentially Important Sources of Variation*	Comments
1	Automated devices calibrated periodically	Common general chemistries and immunoassays; microarrays	• Run • Day • Calibration cycle • Calibrator lot • Reagent lot • Microarray chips and related effects	If a system needs to be recalibrated at relatively long time intervals (eg, 6 months or more), then inclusion of different calibration cycles in the precision study may not be necessary.
2	Automated devices calibrated frequently (per run or per day)	Selective electrode chemistries (Na, K, Cl, CO_2)	• Run • Day • Calibration cycle • Calibrator lot • Reagent lot • Microarray chips and related effects	Because a system is calibrated at each run (or day), calibration variability is included in between-run component (or between-day component).
3	Automated devices that are not calibrated	Enzyme assays without calibrators (eg, amylase, CK, LDH on multianalyte systems)	• Run • Day • Reagent lot	
4	Automated devices with factory-assigned calibration for each reagent lot	Some molecular diagnostic virology and microbiology assays	• Run • Day • Reagent lot	Calibrator lot-to-lot variability is included in between-lot component.
5	Automated devices with factory-assigned calibration for each reagent lot and manual or semiautomated pretreatment step(s)	Some molecular diagnostic virology and microbiology assays	• Run • Day • Reagent lot • Operator	Calibrator lot-to-lot variability is included in between-lot component.
6	Manual or semiautomated calibrated devices	Lateral flow assays requiring pretreatment and operator interpretation; classical RIAs and ELISA plate methods	• Run • Day • Operator	
7	Measurement procedure always used only within a single laboratory or laboratory system	Assays developed by a clinical laboratory for its own use	• Run • Day • Calibration cycle • Calibrator lot • Reagent lot • Operator (if manual or semiautomated)	• Operator is important if the assay is manual or semiautomated. • Reagent lot is important when reagent lots are laboratory developed.

Abbreviations: CK, creatine kinase; ELISA, enzyme-linked immunosorbent assay; LDH, lactate dehydrogenase; RIA, radioimmunoassay.

Depending on the particular device, the sources of variation listed here may not all be relevant and appropriate to incorporate in the experiment, whereas other sources of variation not listed may be relevant and should be included.

© Clinical and Laboratory Standards Institute. All rights reserved.

2.3 Suggested Study Designs

This section briefly introduces the two standardized precision evaluation study designs described more fully later in this document.

2.3.1 Single-Site Study: the Classic 20 × 2 × 2 Design

The single-site study is the original EP05 precision protocol and remains well suited for many types of measurement procedures in the clinical laboratory. The 20 × 2 × 2 nomenclature comes from the basic design: test each sample twice per run, two runs per day, for 20 or more testing days (not necessarily consecutive) using a single instrument at a single site. Use of this protocol provides estimates of repeatability and within-laboratory precision. See Chapter 3 for a detailed account of this protocol.

NOTE:

See Chapter 3 for a detailed description of the **classic precision protocol.**

2.3.2 Multisite Study

The multisite study protocol provides estimates of reproducibility, as well as repeatability and within-laboratory precision. It consists of replicating across "sites" (meaning laboratories and/or instruments) an experimental design capable of yielding precision estimates comparable to those featured in the single-site study. At minimum, for each sample, it should involve three sites, five (not necessarily consecutive) days, and 20 *DF* per site for repeatability, which can be achieved, for example, by assaying five replicates per run, with one run per day, or by assaying three replicates per run, with two runs per day. The protocol may be appropriate in cases where there is clinical relevance to between-site measurements, as well as for the introduction of new markers and/or measurement technologies. A reproducibility study by the manufacturer of a new assay can help to demonstrate robustness of the assay in the context of the wider community of end-user clinical laboratories. See Chapter 4 for a detailed account of the 3 × 5 × 5 protocol.

A **multifactor study** is appropriate when multiple sources of variation, which may include multiple sites, are to be considered explicitly. Consultation with a statistician is suggested to ensure that all experimental factors and levels of the design are properly accounted for in the analysis model. See Appendix C for an extensive discussion of **multifactor study protocols,** including examples of models optimized for particular types of measurement procedures.

2.4 Multifactor Studies

A multifactor study is appropriate when multiple sources of variation, which may include multiple sites, are to be considered explicitly. Consultation with a statistician is suggested to ensure that all experimental factors and levels of the design are properly accounted for in the analysis model. See Appendix C for an extensive discussion of multifactor study protocols, including examples of models optimized for particular types of measurement procedures.

© Clinical and Laboratory Standards Institute. All rights reserved.

This page is intentionally left blank.

© Clinical and Laboratory Standards Institute. All rights reserved.

Chapter 3

Single-Site Precision Evaluation Study

This chapter includes:

- Overview, general guidelines, and factors for consideration when performing a single-site precision evaluation study
- Recommended protocol design and requirements for a single-site precision evaluation study
- Experimental steps and data analysis for single-site precision study

Single-Site Precision Evaluation Study

3.1 Introduction

The single-site precision evaluation protocol discussed here provides a means to establish, for a quantitative measurement procedure, estimates of repeatability and within-laboratory precision that reflect the performance a user can expect to achieve in routine practice with a single instrument in a single laboratory using a single reagent lot over the course of a month or so, if the manufacturer's instructions are followed. The resulting precision estimates may be verified by users following CLSI document EP15.[1]

The single-site protocol is eminently suitable when prior assessment of the assay (see Chapter 2) indicates that variability from sources such as reagent lot-to-lot, instrument-to-instrument, calibration cycle-to-cycle, and operator-to-operator variability are minimal. Appendix C provides guidance on more comprehensive evaluation protocols for situations where, in assessing an assay's precision characteristics, one or more of these or other sources must be taken explicitly into account, ie, treated as factors in the statistical model.

The **single-site protocol** discussed in this section follows the general approach to within-laboratory precision studies described in ISO 5725-5.[26]

In the context of validating a new assay, the presumption that reagent lot and instrument sources of variation are relatively small compared to other sources may not apply, and yet there will usually not be a sufficient number of reagent lots or instruments available for testing to yield a solid assessment of the impact of these potential sources of variation on the assay's precision performance characteristics. Accordingly, it is common practice to perform the 20 × 2 × 2 experiment described in this section using just one reagent lot and a single instrument, repeating the entire experiment using the same samples for each of three reagent lots, for example, and (for example) two instruments per lot, thereby yielding for each sample, in this scenario, six datasets of ≥ 80 results each. This allows for comparing, across the reagent lots used in the parallel studies, estimates for the variance components and precision types featured in the 20 × 2 × 2 experiment.

The single-site protocol discussed in this section follows the general approach to within-laboratory precision studies described in ISO 5725-5.[26]

The protocol uses a nested components-of-variance design with 20 testing days, two runs per testing day, and two replicate measurements per run (a 20 × 2 × 2 design) for each sample. A single reagent lot and a single calibrator lot are used, and all testing is performed with a single instrument at a single site. Some things that may influence precision, such as sample preparation, test material stability, carryover, and drift (see CLSI document EP10[25]) are understood to be sources of within-laboratory variability, but their contributions are not estimated separately.

© Clinical and Laboratory Standards Institute. All rights reserved.

3.2 General Guidelines

It is assumed that the manufacturer or developer of the measurement procedure has a good understanding of the procedure, equipment, and reference materials. If testing is performed outside the manufacturer's facility, then all testing personnel must be qualified as competent with the equipment, measurement procedure, and experimental protocol. All results for the precision experiment should be generated in accordance with the manufacturer's operating instructions.

It is further assumed that the device is operating in a stable condition during the study's data collection phase. To justify this assumption, performance should be monitored with QC samples and routine internal QC procedures throughout the evaluation period. At least one QC sample at an appropriate concentration should be included in each run. For new assays or devices, sufficient data should be available to establish that the measurements are in statistical control before performing the experiment, and there should be a process for determining that the device remains in control for all runs in the precision experiment. For guidance on QC practices, refer to CLSI document C24.[20]

3.3 Preliminary Checkout

Before embarking on the 20 × 2 × 2 experiment, it may be advisable to perform a brief unstructured within-run assessment in order to check that the measurement procedure is performing within expectations, at least as far as repeatability is concerned. For example, 20 aliquots of each of one or more appropriate test materials (or a complete "batch," if < 20) can be tested in a single run using the instrument and the specific product lots (reagent, calibrator, generic system reagents, etc.), which will be used in the precision evaluation study. Collect the measurand test results, then compute mean, SD, and %CV for each sample tested. If any of these statistics shows a surprising departure from expectations, the main precision evaluation study should not be started until the concern is resolved. Note that the results generated in this repeatability check are not used in the final analysis of the single-site precision study.

© Clinical and Laboratory Standards Institute. All rights reserved.

3.4 Protocol Design and Requirements

The recommended single-site 20 × 2 × 2 precision evaluation protocol follows a nested linear components-of-variance model involving two factors: "day" and "run," with "run" nested within "day."

$$Y_{ijk} = \mu + D_i + R_{j(i)} + \varepsilon_{k(ij)} \quad (3)$$

for $i = 1$ to n_{day}, $j = 1$ to n_{run}, and $k = 1$ to n_{rep}, where:

Y_{ijk} = observed measurement for the kth replicate in the jth run on the ith day

μ = the mean, estimated as the average of the N measurements

D_i = shift due to day-to-day sources of variability on the ith day ($i = 1, \ldots, n_{day}$)

$R_{j(i)}$ = shift due to run-to-run sources in the jth run of the ith day ($j = 1, \ldots, n_{run}$)

$\varepsilon_{k(ij)}$ = residual, ie, shift due to within-run sources for kth replicate in the jth run on the ith day ($k = 1, \ldots, n_{rep}$)

n_{day} = number of days: usually 20

n_{run} = number of runs per day: usually two

n_{rep} = number of replicates (measurements) per run: usually two

N = total number of measurements expected: usually $n_{day} \times n_{run} \times n_{rep} = 80$

The design is balanced because it specifies the same number of runs for each day, and the same number of replicates for each run. (But n_{run} need not be the same as n_{rep} to warrant characterizing such a design as balanced. See the comb diagram in Figure 3 and the associated text in Section 1.5.4.1.)

3.4.1 Number of Days, Runs, and Replicates

The 20 × 2 × 2 experiment is conducted over at least 20 days (not necessarily consecutive working days). Standard calibration practice frequency per the manufacturer's instructions should be followed throughout the testing period.

There will be two runs per testing day, with a recommendation of a minimum of two hours' separation between runs (if possible). Two replicates per sample are tested per run. (See Section 3.4.5 on aliquots.) Based on this design, a total of at least 80 measurements are expected for each sample (20 or more days × 2 runs per day × 2 results per run).

3.4.2 Number of Instruments and Sites

An individual 20 × 2 × 2 experiment is conducted using a single instrument at a single site.

© Clinical and Laboratory Standards Institute. All rights reserved.

3.4.3 Number of Reagent and Calibrator Lots

An individual 20 × 2 × 2 experiment is conducted using a single reagent lot and a single calibrator lot.

3.4.4 Number of Calibration Cycles

The measurement procedure should be calibrated at the start of the study, on "day one," per the manufacturer's instructions. Subsequent calibrations, if any, should be done at the frequency designated in the manufacturer's PI. As such, the precision evaluation study will be completed for some measurement procedures within a single calibration cycle. Others will experience multiple calibration events, thus incorporating cycle-to-cycle variability into the within-laboratory precision estimation. (In reporting the study, the number of calibration events must be stated to make it clear whether—and if so, how well—variability associated with calibration events is accounted for in the within-laboratory precision estimates.)

3.4.5 Number and Types of Samples

Matrix—Test samples for this protocol should be selected to reflect the characteristics of actual samples that would routinely be presented to a user's laboratory for testing. These include primarily patient specimens (more typically: pooled patient specimens) of the kind (eg, serum) that the assay is intended to measure, and secondarily control materials and/or calibrators suitable for routine QC of the assay. It may be necessary to spike or dilute patient samples in order to achieve measurand levels in the high and low regions of the measuring interval where native patient samples cannot be sourced. Such contrived samples should be shown to be commutable with native patient samples. When necessary, stable, commercially available materials with a suitable matrix (eg, protein-based, for assays intended for serum samples) may be used. Matrix effects can create biases, though it is generally expected that their effect on random variability is small and can be ignored. (Comparing precision performance of a control material to that of a patient sample pool of essentially the same measurand level may allow for assessing the impact of such matrix effects.)

Measurand Levels—Samples representing a sufficient number of measurand levels must be tested in the study to adequately characterize the measurement procedure's precision over all or most of its stated measuring interval, including points at or near medically important decision levels within it. It is recommended that the samples prepared for use be screened before the start of the main experiment to check that they represent a suitable spectrum of levels. The number and positioning of the levels tested should be sufficient to ensure that users can extract from the tabulation reasonable estimates for repeatability and within-laboratory precision expected at intermediate levels, not just at measurand levels corresponding to the reported estimates.

© Clinical and Laboratory Standards Institute. All rights reserved.

In many cases, five or more levels will be needed to accommodate measurement procedures with wide measuring intervals and multiple medical decision points or important medical levels near the lower limit of quantitation (LoQ). Three levels may suffice for measurement procedures with limited measuring intervals and only one medical decision point or with consistent precision across all or most of the measuring interval. For assays intended for monitoring, concentration levels should be chosen carefully to reflect the intended use. For measurement procedures with medically important points at the assay's LoQ and/or limit of detection (LoD), it is recommended to include in the precision study measurand levels close to these clinically important points and to use numerical values below the lower limit of measuring interval in an estimation of precision (for LoQ) and percent of results "detected" (for LoD) (for details, consult CLSI document EP17[34]). For other measurement procedures, measurand levels for samples in the high- and low-end regions of the measuring interval should be selected such that most of the test results in the study would be expected to fall within the measuring interval of the measurement procedure. As such, it is not appropriate to use a sample with a measurand level too close to the upper limit of the measuring interval (ULMI) or lower limit of the measuring interval (LLMI).

IMPORTANT NOTE:

It is recommended that a **different aliquot** be used for **each replicate or for each run,** depending on the character and clinical use of the measurement procedure.

Aliquots—When appropriate and feasible, it is desirable to aliquot and freeze the test samples. This approach can minimize extraneous sample-related and sample handling–related sources of variation, providing meticulous attention is given to preparing, sealing, labeling, storing, and handling the aliquots, to ensure sample integrity and to minimize the risk of sample misidentifications, mix-ups, leakage, etc. Over the course of the study, follow consistent practices in dealing with the aliquots. It is recommended that a different aliquot be used for each replicate or for each run, depending on the character and clinical use of the measurement procedure.

Using a different aliquot for each replicate ensures that any step, such as a preliminary extraction step, which ought to be considered an integral part of the assay procedure (when applied in routine practice with patient specimens) will be reflected, with substantial *DF*. (Run-to-run, day-to-day, and replicate-to-replicate variation all contribute to within-laboratory precision and give information regarding variation that is especially relevant when patients are monitored or resampled over time. In some settings, multiple samples collected from a patient over a short time frame—eg, before and after a particular stimulation—are all assayed in a single run: in this situation, replication error may be the primary source of imprecision.)

The rationale behind the recommendation involves a balancing of two competing considerations: 1) The precision experiments should reflect sources of variation from all steps in the measurement procedure as it would be applied in clinical practice; and 2) the experiments should be

© *Clinical and Laboratory Standards Institute. All rights reserved.*

designed to minimize the impact of extraneous sources on precision estimates, outlier contamination, etc. For most assays, extraneous sources include the preparation and handling of aliquots, because these processes would not be involved in the routine measurement of patient samples. Moreover, for a measurement procedure involving nothing analogous to a preliminary extraction step, the first consideration simply does not apply.

3.5 Experimental Steps

1. Calibrate the measurement procedure following the manufacturer's instructions at the start of the study. Periodically recalibrate during the study at the frequency designated by the manufacturer.

2. Perform two runs on "day one" of the study, with a recommendation of a minimum of two hours' separation between runs (if possible). In each run, generate two measurements for each test sample. Include QC sample(s) in each run. Randomize the order of samples for each run or day.

   NOTE: See Section 3.4.5 regarding aliquots for guidance on whether or not both measurements may come from the same aliquot.

REMINDER:

See Section 3.4.5 regarding aliquots for guidance on **whether or not both measurements may come from the same aliquot.**

3. Record the measurement obtained for each replicate, if possible using (at least) one more decimal place than would normally be used for reporting patient results, in order to minimize roundoff artifacts in the statistical analysis.

4. Repeat steps 2 and 3 for another 19 days of testing. The days do not have to be consecutive. The time of the runs does not have to be the same each day. Indeed, if possible, the run times should be varied throughout the work day, maintaining at least a two-hour separation between runs (if possible), to introduce more randomness into the testing. It is desirable to randomize across samples the order in which the aliquots are measured—providing that, with due vigilance, this can be done without putting the identity of the results at risk, ie, without causing irreparable sample mix-ups.

5. Review the results of each run promptly to ensure that all tests were completed properly and that the measurement procedure continues to perform under acceptable statistical control. Assess the data for potential outliers, trending, error messages, and atypical results.

   NOTE 1: Any missing results for individual replicates—due, for example, to instrument malfunction, or to unrectifiable operator or data recording errors—should be made up during the given run, if possible, and any missing runs should be repeated, preferably the same morning or afternoon. If that is not possible, another full

© Clinical and Laboratory Standards Institute. All rights reserved.

day should be added to the study. For each test sample, a dataset of at least 80 measurements (representing 20 days or more) must be available for analysis at the end a 20 × 2 × 2 study.

NOTE 2: All values reported by the instrument are to be used unless specifically flagged with an error message that would preclude normal reporting, or if an error condition is identified that would explain an atypical result (eg, sample mishandling). **Do not** exclude any results from the final dataset without documented cause.

NOTE 3: If a run must be rejected because of QC procedures (invalid run), conduct an additional run on the same day, if possible, after an investigation is conducted to identify and correct the cause of the problem. Maintain a record of the number of invalid runs. If more than 5% of the runs need to be rejected and no assignable cause can be found, then the investigator should consider the possibility that the device is not sufficiently stable to allow for a valid assessment of its precision. It may be desirable to schedule additional days of evaluation at the outset of the investigation, to allow for potential run rejections, if needed.

3.6 Data Analysis

For ease of understanding, this section assumes that data analyses will be applied to balanced datasets. The text presents basic ANOVA calculations and interpretations. If the dataset for one or more of the test samples is unbalanced, it will be necessary to use a more advanced data analysis technique, as noted in Section 3.9.

3.6.1 Checking Data Integrity

Throughout the study, review the data periodically, preferably in real time, to ensure that measurements are being obtained as expected, the values appear reasonable, and the assumptions of the experiment are being met. Any suspect results should be identified as soon as possible to allow prompt follow-up investigation and, if warranted, repeat testing. Waiting until the end of the study to review the data may necessitate repeating the entire experiment for one or more samples if it is no longer possible to determine the assignable causes for highly discrepant measurements.

The final dataset for each sample should be reviewed once more after completion of the benchwork phase. Histograms, trend plots, and other displays (see examples in Appendixes A and B) may be useful for inspecting the data at this stage. Individual measurements or even clusters of measurements that seemed acceptable while the study was in progress might now stand revealed as potential outliers against the full body of results. They should be double-checked for transcription errors, sample mix-ups, or other anomalies. Carefully document all corrections and exclusions for assignable cause.

In the final analysis, to do justice to what the bulk of the data indicates about the assay's precision, it may be desirable to treat certain highly

© Clinical and Laboratory Standards Institute. All rights reserved.

discordant measurements as "statistical outliers." This is acceptable providing that three conditions are met:

The discordancy of each result treated as a statistical outlier must be documented via an objective statistical outlier test.[27]

NOTE: Even if a result qualifies in this manner as highly aberrant, it is not necessary to treat it as a statistical outlier. Doing so may not have any substantial impact on the precision estimates, and the fewer results treated as outliers the better.

A. In general, for any single block of data—consisting of the ≥ 80 results for a 20×2×2 or 20×1×4 experiment (or the ≥ 60 results for a 20×1×3 experiment)—no more than **two** results should be treated as statistical outliers.

B. For the multisample study as a whole, the individual measurements treated as outliers should constitute no more than **1%** of the totality of results.

NOTE: Jointly, these two clauses allow for treating a pair of results (possibly representing an entire run) as statistical outliers for **any one** sample, but not for **every** sample in the study. In short, clause 2a applies locally, and clause 2b applies globally. For a multisample precision study, and more generally for a family of such studies (each representing, for example, a different reagent lot/instrument combination), which could well entail several thousand results altogether, the overall distribution of highly aberrant results is expected to be such that, at most, a few blocks of data will be affected, while many will not.

Any result treated as a statistical outlier belongs to the final dataset. Accordingly, data analysis for the precision study must be performed and summarized twice, ie, **both before and after** suppressing these discordant results, to demonstrate their impact on the precision estimates. The percentage of outliers should be reported; and their clinical significance should be evaluated.

3.6.2 Precision Estimates Calculated by Analysis of Variance

Estimates for the variance components and for the precision types of interest—in this experiment: repeatability and within-laboratory precision—are calculated one sample at a time, desirably with the help of existing software that is well documented and suitably vetted. (The sample-specific statistics should then be assessed in aggregate for what they have to say about the assay's precision characteristics across all or most of its measuring interval—minimally via a summary table, and desirably via precision profiles. See Sections 3.8.1 and 3.7.2, respectively.)

The 20×2×2 design specified in Section 3.4 is fully balanced, and it sets expectations for a total of exactly 80 results per sample. The final dataset, however, may contain more than 80 results if the study is extended

© Clinical and Laboratory Standards Institute. All rights reserved.

NOTE:

For more information on **unbalanced ANOVA** and alternative methods of data analysis, **consult a statistician,** the references in this guideline, or the documentation for suitably vetted software.

NOTE:

Terminology and layout for the standard **two-way nested ANOVA** summary table may vary. Table 2 illustrates the "σ^2" notation for variances rather than, for example, "*V*" or "*Var.*"

beyond 20 days, and it may be unbalanced if there are missing values or if any results are treated as statistical outliers in the final analysis.

To keep the exposition simple, Sections 3.6.2 and 3.6.3 assume that the dataset to be analyzed mirrors the 20 × 2 × 2 experimental design in consisting of the 80 measurements expected per sample. This makes the dataset amenable to analysis via a balanced two-way nested ANOVA routine (based on the "method of moments" wherein variance components are estimated by matching expected mean squares [EMS]), which is fairly straightforward to explain and interpret, and for which software implementations are widely available. The same approach can be used for datasets with no missing values from experiments involving additional (or fewer) days, that is, any $n_{day} \times 2 \times 2$ design, or more generally any balanced two-way design, such as the 3 × 5 × 5 reproducibility study discussed in Chapter 4.

When there are missing values or measurements set aside as statistical outliers, this approach is no longer viable, but the data can be analyzed by a slightly more complicated two-way nested ANOVA routine capable of dealing with unbalanced datasets alternatively, by REML, MINQUE, MIVQUE, or various related methodologies.[23] These routines, being suitable for use on balanced as well as unbalanced datasets, have greater applicability than the ANOVA routine described herein; moreover, on balanced datasets, they can be expected to yield essentially the same statistical results. For more information on unbalanced ANOVA and alternative methods of data analysis, consult a statistician, the references in this guideline, or the documentation for suitably vetted software.

Table 2 displays a schematic ANOVA table representing, for any one sample, the pivotal output generated by most methods for fitting the two-factor, nested random-effects model formulated in Section 3.4 to a complete dataset (no missing values) arising from an experiment conforming to the 20 × 2 × 2 design. The specific values in the *DF* column are based on the recommended dimensions for the single-site study, namely, n_{day} = 20 days, n_{run} = 2 runs per day, and n_{rep} = 2 replicates per run, implying a total of N = 80 results per sample.

Table 2. Generic Two-Way Nested ANOVA Summary Table Format, for Single-Site 20 × 2 × 2

Source of Variation	*SS*	*DF*	*MS*	EMS
Day	SS_{day}	$DF_{day} = n_{day} - 1 = 19$	MS_{day}	$\sigma^2_{error} + n_{rep}\sigma^2_{run} + n_{run}n_{rep}\sigma^2_{day}$
Run (Day)	SS_{run}	$DF_{run} = (n_{run} - 1)n_{day} = 20$	MS_{run}	$\sigma^2_{error} + n_{rep}\sigma^2_{run}$
Error	SS_{error}	$DF_{error} = N - n_{day}n_{run} = 40$	MS_{error}	σ^2_{error}
Total	SS_{total}	$DF_{total} = N - 1 = 79$		

Abbreviations: *DF*, degrees of freedom; EMS, expected value of the mean square; *MS*, mean squares; *SS*, sum of squares.

© Clinical and Laboratory Standards Institute. All rights reserved.

Terminology and layout for the standard two-way nested ANOVA summary table may vary. Table 2 illustrates the "σ^2" notation for variances rather than, for example, "*V*" or "*Var*." Often omitted is the EMS column, which documents how the *MS* values are to be interpreted as functions of the three variance components and the *n*-sizes.

The *DF* (or *df*) for each source of variation can be determined directly (as noted in the *DF* column) from the *n*-sizes (number of days, number of runs per day, and number of replicates per run) associated with the 20 × 2 × 2 design, because a complete balanced dataset is assumed here for the sake of illustration.

NOTE 1: Table 2 lists the "total" *DF,* here designated DF_{total}, as N – 1. This corresponds to the *DF* that would apply to a set of N results exhibiting no structure, ie, when just one uniform source of variation affects the results. When ANOVA is appropriate due to the presence of more than one source of variation, the *DF*, df_{WL}, approximated (usually by the Satterthwaite method)[22] for the overall precision estimate, S_{WL}, will always be less than N – 1, and often considerably less.

NOTE 2: The calculations in this section can be readily adapted to handle balanced datasets involving different study dimensions (*n*-sizes) and/or different but similarly nested sources of variation. (In fact, except for different factor names and different *n*-sizes, the calculations described here are essentially the same as those described for a multisite 3 × 5 × 5 dataset in Section 4.6.2.)

If the software does not also report the variance components V_{day}, V_{run}, and V_{error}, calculate them from the ANOVA table's *MS* values and the study's dimensions (*n*-sizes) as follows.

$$V_{error} = MS_{error} \quad (4)$$

$$V_{run} = \text{the maximum of: 0 or } \frac{(MS_{run} - MS_{error})}{n_{rep}} \quad (5)$$

$$V_{day} = \text{the maximum of: 0 or } \frac{(MS_{day} - MS_{run})}{n_{run} n_{rep}} \quad (6)$$

As indicated in equations (5) and (6), "negative variance components" are dealt with in the ANOVA approach by resetting the estimates to zero. (In a REML analysis, they do not arise).[23] Note that a zero estimate for, eg, the day-to-day variance component, V_{day}, does not mean that one will observe no variability from one day to the next, but rather that the within-run and/or run-to-run sources may so dominate the overall variability inherent in the dataset that the day-to-day source appears negligible by comparison.

© Clinical and Laboratory Standards Institute. All rights reserved.

The repeatability SD (S_R) corresponds directly to the residual or error-term variance component (V_{error}), while the within-laboratory precision SD (S_{WL}) corresponds to the sum of all three variance components:

$$S_R = \sqrt{V_{error}} \quad (7)$$

$$S_{WL} = \sqrt{V_{day} + V_{run} + V_{error}} \quad (8)$$

These may be re-expressed as %CVs: simply divide by the grand mean of all measurement results for the sample, then multiply by 100.

NOTE: The value of the overall SD, S_{WL}, as determined from equation (8) can generally be expected to differ somewhat from the value obtained by simply calculating an SD directly from the set of N observations. Usually, however, it is the *DF*, df_{WL}, assigned to the overall SD that constitutes the most salient difference between the direct approach and the variance components (ANOVA) approach. The former assumes, in effect, a single homogeneous source of variation, whereas the latter models overall variability as a combination of multiple (in this design: three) hierarchically related sources.

3.6.3 Confidence Intervals for the Precision Estimates

For a given sample, the calculations in Section 3.6.2 yield estimates for repeatability and for within-laboratory precision. For each of these precision types, from the estimate and its *DF*, it is possible to construct a CI bracketing the estimate (though doing so is generally not required). For practical purposes, this can be thought of as an interval within which we can have a stated level of confidence that the sample's true SD (or %CV) for that precision type lies, or as an interval representing the spectrum of SD (or %CV) values for that precision type that could plausibly be attributed to the sample, in light of the data underpinning the estimate. CIs thus serve to characterize the imprecision or uncertainty of the estimates obtained, primarily with respect to the experiment's size or, more exactly stated, the extent to which the relevant sources of variation are jointly exercised in the study (as discussed in Section 1.5.4.2).

Illustrated below is the chi-square (χ^2)–based approach to constructing CIs, which relies on the original Satterthwaite approximation[22] for df_{WL}, the within-laboratory precision estimate's *DF*. Once again, for simplicity, the illustration is limited to the balanced case involving 80 results per sample. Several other approaches may be used, either for approximating df_{WL} or for constructing the CIs. For these, consult a statistician, the references in this guideline,[28-30] or the documentation of suitably vetted software. Be mindful that the various approaches and software implementations can be expected to differ somewhat in their df_{WL} approximations and in the confidence limits constructed for the within-laboratory precision estimate.

Assuming a balanced design, for a complete dataset (no missing values), the repeatability estimate's *DF* can be identified with the error term's *DF* recorded in the ANOVA table:

$$df_R = DF_{error} = \mathrm{N} - n_{day} n_{run} \quad (9)$$

© Clinical and Laboratory Standards Institute. All rights reserved.

The within-laboratory precision estimate's *DF*, on the other hand, cannot be identified with the ANOVA summary's DF_{total} (see the bottom-most line in Table 2), which represents the *DF* that would apply if there were just one, homogeneous source of variation. It is important to understand that df_{WL}, the statistic relevant here, is always less than DF_{total}, and usually considerably less. Accordingly, if DF_{total} (80 – 1 = 79 for an 80-measurement 20 × 2 × 2 dataset) were to be used, mistakenly, in constructing a CI for S_{WL}, the resulting interval, being too narrow, would constitute an unduly optimistic assessment of the within-laboratory precision estimate.

Because df_{WL} represents the *DF* for a linear combination of the variance component estimates and the residual variability, it depends on both the relative magnitude of the three sources and the extent to which they are exercised in the study. Accordingly, df_{WL} must be approximated as a fairly complex function of both the *MS* and the *DF* tabulated for day, run, and residual error. The original Satterthwaite approximation[22] is widely used for this purpose, though it may not be optimal.[28-30] (In practice, which method is used for the df_{WL} approximation may already have been determined by the software chosen for generating the precision estimates.) Equation (10) shows the Satterthwaite formula for complete datasets arising from a balanced single-site study:

$$df_{WL} = \frac{(\alpha_{day} MS_{day} + \alpha_{run} MS_{run} + \alpha_{error} MS_{error})^2}{\frac{(\alpha_{day} MS_{day})^2}{DF_{day}} + \frac{(\alpha_{run} MS_{run})^2}{DF_{run}} + \frac{(\alpha_{error} MS_{error})^2}{DF_{error}}} \quad (10)$$

The *DF* terms on the right of the equal sign come from the ANOVA summary (see Table 2), as do the *MS*. The alpha coefficients are $\alpha_{error} = 0.50$, $\alpha_{run} = 0.25$, and $\alpha_{day} = 0.25$ for a complete (balanced, 80-result) 20 × 2 × 2 dataset. If the dataset's dimensions (*n*-sizes) differ from this norm, as they often will, consult the literature[23,31-33] (or a statistician, or rely on suitably vetted software). Note that the Satterthwaite approximation must be expected to yield a numerical result with a fractional component, rather than a simple positive integer (eg, 23.45 as opposed to 23).

When evaluating or validating precision performance, we are generally interested in 95% CIs, corresponding to an "alpha" level of 5% ($\alpha = 0.05$). In that case, at a CL of $1 - \alpha = 0.95$, the limits of a precision estimate's CI are calculated as:

$$S\sqrt{\frac{DF}{\chi^2_{(1-\frac{\alpha}{2}),DF}}} \quad \text{and} \quad S\sqrt{\frac{DF}{\chi^2_{\frac{\alpha}{2},DF}}} \quad (11)$$

where: S = the precision estimate, expressed as an SD

DF = the precision estimate's degrees of freedom

$\chi^2_{CL,DF}$ = chi-square distribution value for the desired CL and DF

NOTE:

The pair of terms in equation (11) represents a **two-sided interval.** For the usual choice of $\alpha = 0.05$, this corresponds to a lower limit at $\chi^2_{0.975,DF}$ and an upper limit at $\chi^2_{0.025,DF}$.

© Clinical and Laboratory Standards Institute. All rights reserved.

NOTE: The pair of terms in equation (11) represents a two-sided interval. For the usual choice of $\alpha = 0.05$, this corresponds to a lower limit at $\chi^2_{0.975,DF}$ and an upper limit at $\chi^2_{0.025,DF}$. In some contexts, where only the upper limit is of interest, a "one-sided interval" may be more appropriate. For $\alpha = 0.05$, this corresponds to an upper confidence limit at $\chi^2_{0.05,DF}$.

Having constructed 95% CIs for the precision estimates expressed as SDs, simply divide these limits by the mean, and multiply by 100, to obtain **approximate** 95% CIs for the precision estimates expressed as %CVs. Providing the %CVs are less than 30% or 40%, this method of constructing CIs for them represents a satisfactory approach to characterizing their uncertainty as far as sampling variation is concerned.

NOTE 1: Neglecting to account for uncertainty in the mean, the approximation described above fails at very high %CVs, but remains quite accurate up to and well beyond %CVs that are typically used to define an LoQ based on precision requirements, ie, 20% or less for most immunoassays. The coverage of CIs for %CVs, when the CIs are constructed in this manner, is expected to degrade systematically (ie, to fall short of the intended ["nominal"] 95% coverage) essentially as a function of increasing %CV, and it is little influenced by *DF* or N-size. For practical purposes, however, the degradation is negligible at modest %CVs. Thus, for %CVs of 10%, 20%, 30%, and even 40%, the expected coverage remains high: approximately 95%, 94%, 93%, and 91%, respectively. Accordingly, for precision studies restricted to measurand levels above the assay's lower LoQ (see CLSI document EP17[34]) the approach described above can generally be recommended for yielding very good approximations, as well as for its computational simplicity.

NOTE 2: Definitive approaches to constructing CIs for %CVs depend on the noncentral t-distribution,[35] which is not widely available in software for many users. Other relatively complex but still computationally feasible approximations can cope with high %CVs, but only when the dataset is sufficiently large. For example, coverage for nominal 95% CIs constructed by Hald's symmetric method[36] degrades systematically, essentially as a function of decreasing *DF* or N-size, and is little influenced by %CV, achieving an expected coverage of approximately 90%, 92%, 94%, and 94.5% for 3, 10, 30, and 50 *DF*, respectively. In the present context, therefore, these alternative approaches offer no practical advantage over the simple approach recommended above.

3.7 Combining Precision Study Results

Here we sketch two ways of combining the statistics from multiple precision studies: "pooling" and profiles. In some situations, these may yield additional insight into the performance characteristics of a measurement procedure.

© Clinical and Laboratory Standards Institute. All rights reserved.

3.7.1 "Pooling"

Manufacturers often conduct multiple single-site precision studies as described above for each of several different reagent lots, instruments, etc., using a common set of samples. Each study must be fully analyzed and summarized on its own; but in addition, as long as certain assumptions are met, estimates for a given sample and a given precision type (repeatability within-laboratory precision) may be combined ("pooled") across the studies, as explained below, to yield more definitive estimates. (Something similar is done in meta-analysis.[37,38] See also CLSI Document EP17.[34]

Assumptions under which it is possible to combine precision estimates include the following and, in some circumstances, other conditions as well:

- Common sample handling practices and performance of the assay were used in all the studies.
- The data come from studies having essentially the same variance but possibly slightly different means. (A statistical test for homogeneity of variance may be needed to assess this assumption.)
- The distribution of data for each study is approximately gaussian.

The pooled precision estimate, S_{pool}, for a given sample and precision type across multiple, single-site precision studies is calculated as a certain weighted average, shown below. (The same equation serves for both repeatability and within-laboratory precision estimates.)

$$S_{pool} \sqrt{\frac{\sum_{i=1}^{k} DF_i s_i^2}{\sum_{i=1}^{k} DF_i}} \quad (12)$$

where:

s_{pool}	= pooled precision estimate (expressed as an SD)
s_i	= precision estimate (expressed as an SD) from the ith study
DF_i	= DF for the precision estimate from the ith study
k	= number of separate precision studies

The *DF* associated with the pooled estimate equals the sum of the *DF* for the precision estimates from the individual studies.

REMINDER:

A **profile** is thus a supplementary way of summarizing aspects of the statistical analyses. Examples are given in Appendix B. See also CLSI documents EP29[21] and EP17.[34]

© Clinical and Laboratory Standards Institute. All rights reserved.

3.7.2 Precision Profiles

The approach described above in Section 3.7.1 does not provide for combining precision estimates across samples with different measurand levels. However, whether or not the same samples figure in the different studies, fitting a variance function by a suitable curve-fitting technique to the means and variance estimates obtained, and using the variance function to construct a precision profile, represents another useful way of combining information across samples and measurand levels (within or across studies) for a given precision type, providing the sample means represent a suitably broad spectrum of values and assumptions of the regression analysis are met. A precision profile is a graphical representation of precision as a continuous function of concentration, usually displayed with precision expressed in terms of %CV or SD on the vertical axis, and concentration on the horizontal axis.[39,40] It can be further enhanced by adding a confidence band (CB) for the fitted curve.

A profile is thus a **supplementary** way of summarizing aspects of the statistical analyses. Examples are given in Appendix B. See also CLSI documents EP29[21] and EP17.[34]

3.8 Suggested Format for Summarizing Single-Site Precision Study Results

3.8.1 Summarizing for a Package Insert and for Submission

Product labeling for the precision study results should contain a statement describing the evaluation protocol used. In particular, the PI should include all of the study's relevant elements, such as the number of days, runs per day, and replicates per run; the number of calibration cycles, instruments, operators, reagent lots, and calibrator lots; and for each sample, the total number of individual measurements; the number of missing values, invalid runs, invalid results, and results treated as statistical outliers.

Summaries of precision performance characteristics based on a single-site study must include at least the following information for each sample tested:

- Sample description (eg, pooled patient specimens, QC material)
- Mean measurand level
- Number of measurements
- Estimate of repeatability, expressed both as SD and as %CV
- Estimate of within-laboratory precision, expressed both as SD and as %CV

A useful means to organize these basic statistics in a PI is through a summary table like the following, desirably with samples listed in ascending order by measurand level. (Other formats are also in common use.)

Estimates for these variance components are **standard output for two-way ANOVA software.** (See Section 3.6.2, Table 2.)

© Clinical and Laboratory Standards Institute. All rights reserved.

Sample Description	Mean Value	Repeatability		Within-Laboratory Precision	
		SD	%CV	SD	%CV

Abbreviations: SD, standard deviation; %CV, coefficient of variation expressed as a percentage.

Regulatory agencies may require a more complete statistical summary of the data analysis in submissions and possibly in PI tables as well—including, specifically, estimates expressed as SD and %CV for the between-run (within-day) and between-day variance components, as illustrated in the following table. Estimates for these variance components are standard output for two-way ANOVA software. (See Section 3.6.2, Table 2.)

Sample Description	Mean Value	N	Repeatability		Between-Run		Between-Day		Within-Laboratory	
			SD	%CV	SD	%CV	SD	%CV	SD	%CV

Abbreviations: N, number of results; SD, standard deviation; %CV, coefficient of variation expressed as a percentage.

Optional Information. Reporting some or all of the following information is optional, but often highly informative:

- *DF* for the repeatability and within-laboratory precision estimates
- CIs for the repeatability and within-laboratory precision estimates
- Precision profiles delineating the estimated relationship between precision and measurand level for repeatability and/or within-run precision (see Appendix B for examples)
- CBs for the precision profiles
- Diagnostic plots of the measurements in order of generation, or by day and run (see, for example, Figure A1 in Appendix A, and Figures B1 and B2 in Appendix B)

3.9 Note on Unbalanced Datasets

As explained in Section 1.5.4.1 (in connection with Figure 3), the single-site 20 × 2 × 2 precision study and the multisite 3 × 5 × 5 reproducibility study represent fully balanced experimental designs, which can therefore be expected to yield for each sample a fully balanced dataset for analysis, providing that no results are missing or treated as statistical outliers. To

© Clinical and Laboratory Standards Institute. All rights reserved.

keep the exposition simple, the discussion has made this assumption, and for the same reason, it has illustrated just one (widely adopted, conventional) approach to constructing CIs. For an unbalanced dataset—where, for example, not all runs have the same number of results or not all days have the same number of runs—a similar but somewhat different analysis applies. In particular, calculations differ for the EMS coefficients and for the *DF* displayed above in the ANOVA table and subsequent formulae.[31-33] Moreover, for constructing CIs, there are acceptable (arguably superior) alternatives to the Satterthwaite/chi-square approach.[28-30, 41] Software applications are available for generating the required precision estimates and CIs from unbalanced as well as balanced datasets.

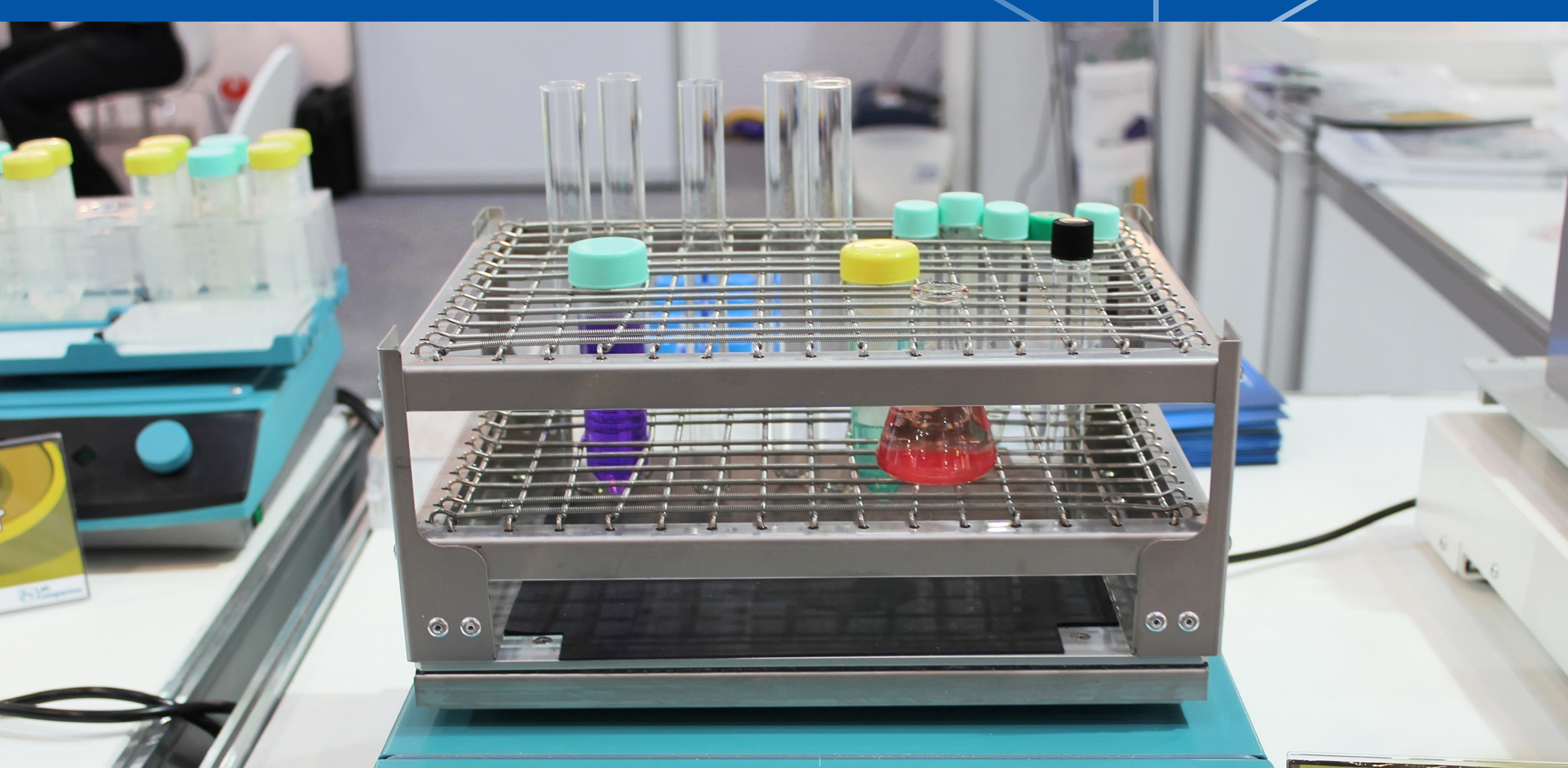
© Clinical and Laboratory Standards Institute. All rights reserved.

Chapter 4

Multisite Precision Evaluation Study

This chapter includes:

- Overview, general guidelines, and factors for consideration when performing a multisite precision evaluation study
- Recommended protocol design and requirements for a multisite precision evaluation study



Multisite Precision Evaluation Study

4.1 Overview

The multisite precision evaluation protocol is intended for use in cases where an assessment of the assay (see Chapter 2) indicates that site-to-site variability may be substantial (as expected for manual or semiautomated measurement procedures), and for situations involving a relatively new measurand. The protocol is meant to complement—**not replace**—the single-site protocol (see Chapter 3) and may be performed concurrently with that evaluation, or subsequently.

If the preliminary assessment indicates that other sources of variation beyond site may be important (eg, reagent lot-to-lot, calibration cycle-to-cycle), one of the more comprehensive multifactor protocols described in Appendix C may be more appropriate to use for precision evaluation. It is advisable to consult with a statistician and the relevant regulatory agencies for guidance in such cases.

It is advisable to **consult with a statistician and the relevant regulatory agencies for guidance** when using comprehensive multifactor protocols, such as those described in Appendix C.

The multisite experiment has a nested components-of-variance design. It requires a minimum of three sites: three different laboratories or, in some cases, three instruments (see Section 4.2). All sites use the same samples. At each site, the samples are assayed on each of five or more days, one run per day, five replicates per run. Overall, this constitutes a 3 × 5 × 5 design, which may also be regarded as a system of three parallel 5 × 5 designs.

This provides a standardized approach to generating site-specific estimates of repeatability and within-laboratory precision at multiple sites, by replicating across the participating sites an experimental design similar to but more compact than the single-site protocol described in Chapter 3. Compared to the 20 × 2 × 2 design, the within-site experiments can each be completed in as few as five days (as opposed to 20). And, because they involve just one factor (rather than two), the precision estimates can be extracted from the data generated at each site using one-way (instead of two-way) ANOVA.

The multisite protocol has two main purposes:

- It allows for comparing site-specific estimates of repeatability and within-laboratory precision for consistency across the participating sites. Because the protocol calls for using the same samples at all the sites, these estimates can also be readily combined, into more definitive estimates of the two precision types.
- It also allows for extracting an estimate of reproducibility from the combined dataset, using two-way nested ANOVA, treating "sites" and "days" as the two factors.

A suitable alternative to the multisite 3 × 5 × 5 design consists of three sites, five days, two runs per day, and three replicates per run—a 3 × 5 × 2 × 3 design that entails a somewhat more complex statistical model (and data processing by three-way nested ANOVA), but likewise yields for each sample 20 *DF* per site for the site-specific repeatability estimates.

© Clinical and Laboratory Standards Institute. All rights reserved.

4.2 General Guidelines

It is assumed that the manufacturer or other developer of a measurement procedure has a good understanding of the procedure, equipment, and reference materials. For testing done outside the manufacturer's facility, all testing personnel must be qualified as competent with the equipment, measurement procedure, and experimental protocol. All results for the multisite precision experiment should be generated in accordance with the manufacturer's PI.

In general, this protocol is expected to be performed across at least three laboratories, each using a single instrument. Two sites should be external and the third site may be internal, relative to the manufacturer or developer. It may be appropriate in some cases to conduct the protocol across three separate instruments within the same laboratory (internal or external). This may be particularly suitable when the device is minimally susceptible to environmental conditions and operator skill such that these contributions to variability may be considered negligible. It is suggested to consult with relevant regulatory agencies before deciding what to designate as the sites (independent laboratories vs different instruments in a single laboratory).

It is assumed that the devices are operating in stable conditions while generating the study data. To justify this assumption, performance should be monitored with QC samples and routine QC procedures throughout the evaluation period. At least one QC sample at an appropriate concentration should be included in each run. For new ssays or devices, sufficient data should be available to establish that the measurements are in statistical control before running the experiment and to establish a process for determining that the device remains in control for all runs in the precision experiment. For guidance on QC practices, see CLSI document C24.[20]

NOTE:

It is assumed that the manufacturer or other developer of a measurement procedure has a good understanding of the procedure, equipment, and reference materials. For testing done **outside the manufacturer's facility, all testing personnel must be qualified as competent with the equipment, measurement procedure, and experimental protocol.** All results for the multisite precision experiment should be generated in accordance with the manufacturer's PI.

IMPORTANT NOTE:

It is assumed that the devices are operating in stable conditions while generating the study data. To justify this assumption, performance should be monitored with **QC samples and routine QC procedures** throughout the evaluation period.

REMINDER:

For guidance on QC practices, see CLSI document C24.[20]

4.3 Preliminary Checkout

It is suggested that a preliminary assessment of repeatability be performed at each site before starting the formal precision evaluation protocol to ensure that the measurement procedure is performing at the sites within the manufacturer's expectations. Twenty aliquots of an appropriate test material (or a complete "batch" if < 20) can be tested in a single run using the instrument and specific product lots (reagent, calibrator, generic system reagents, etc.) which will be used in the precision evaluation study. Collect the analyte test results and compute the site-specific mean, SD, and %CV for each sample tested at each site. If any of these statistics shows a surprising departure from expectations, no further testing should be done until the concern is resolved. Ideally, this preliminary study should be performed for the same measurand levels at all sites.

© Clinical and Laboratory Standards Institute. All rights reserved.

4.4 Protocol Design and Requirements

The recommended 3 × 5 × 5 multisite precision evaluation protocol follows a nested, linear components-of-variance model involving two factors: "site" and "day," with "day" nested within "site."

$$Y_{ijk} = \mu + S_i + D_{j(i)} + \varepsilon_{k(ij)} \quad (13)$$

for $i = 1$ to n_{site}, $j = 1$ to n_{day}, and $k = 1$ to n_{rep}, where:

Y_{ijk} = observed measurement for the kth replicate on the jth day at the ith site

μ = the mean, estimated as the average of the N measurements

S_i = shift due to sources of variability at the ith site ($i = 1, 2, \ldots, n_{site}$)

$D_{j(i)}$ = shift due to day-to-day sources of variability on the ith day ($i = 1, \ldots, n_{day}$)

$\varepsilon_{k(ij)}$ = residual, ie, shift due to within-run sources for kth replicate in the jth run, on the ith day ($k = 1, \ldots, n_{rep}$)

n_{site} = number of sites: usually three

n_{day} = number of days at each site: usually five

n_{rep} = number of replicates (measurements) per day: usually five

N = total number of measurements expected per sample: usually $n_{site} \times n_{day} \times n_{rep} = 75$

The design is balanced because it specifies the same number of days for each site, and the same number of replicates for each day, but n_{day} need not be the same as n_{rep}. See the second comb diagram in Figure 3 in Section 1.5.4.1.

Days and runs are confounded in the 3 × 5 × 5 design because only one run is performed each day. The 3 × 5 × 2 × 3 variant of this multisite protocol avoids that confounding by incorporating three factors in the model: sites, days (nested within sites), and runs (nested within days). The 3 × 5 × 5 and 3 × 5 × 2 × 3 designs both allow for estimating the three precision types of interest. Both are balanced designs involving three sites and five testing days, and both imply the same *DF* for repeatability. The 3 × 5 × 2 × 3 variant requires more benchwork and a (somewhat more complex) three-way nested ANOVA or some equivalent data processing technique. Its main advantage, if any, lies not in its overcoming the confounding of days and runs, but rather in its somewhat larger size, which should translate into somewhat narrower CIs for the within-laboratory precision and reproducibility estimates, ie, estimates associated with less uncertainty.

NOTE: It is possible to perform the multisite evaluation by adopting at one site the 20 × 2 × 2 protocol (described in Chapter 3) or a suitable variant thereof, while adopting at two other sites a 5 × 5 (five days, five replicates per day) or 5 × 2 × 3 (five days, two runs per day, three replicates per run) protocol. This and other reasonable combinations yield unbalanced experimental designs and datasets, thus complicating the analysis.

© Clinical and Laboratory Standards Institute. All rights reserved.

However, they can still be analyzed using ANOVA and comparable routines available in many statistical software applications. For such an approach, consult with a statistician as needed; the relevant data analysis is not treated here.

4.4.1 Number of Days, Runs, and Replicates

The study is to be conducted over a minimum of five (not necessarily consecutive calendar or working) days. The start dates need not be the same for all sites.

There is one run per testing day, with five replicates of each sample tested per run—alternatively, two runs per day, three replicates per run. (See Sections 3.4.5 and 4.4.5 on aliquots.)

4.4.2 Number of Instruments and Sites

The study is conducted with a single instrument at each of at least three sites. Two sites are external and one may be internal, relative to the manufacturer or developer. An alternative implementation of the model (see Section 4.2) uses three instruments at a single site (internal or external).

4.4.3 Number of Reagent and Calibration Material Lots

As a general rule, conduct the study using the same reagent lot and the same calibrator lot at each site. For the evaluation of a new technology, all sites must use the same reagent and calibrator lots. For other situations, although it would be permissible to use different reagent lots and/or different calibrator lots across the sites, this is strongly discouraged because variability would likely be increased, and lot-to-lot variability would be confounded with site-to-site variability. Use of an advanced model (see Appendix C) would allow explicit incorporation of such factors, if desired.

4.4.4 Number of Calibration Cycles

The measurement procedure should be calibrated at the start of the study at each site, per the manufacturer's instructions. Subsequent calibrations, if any, should be done at the frequency designated in the manufacturer's PI.

4.4.5 Number and Types of Samples

The standard multisite protocol presented in this chapter requires that the same samples be tested at each of the sites. This is often done by providing frozen aliquots of a sample set to each of the sites, assuming that the measurand is tolerant to freeze/thaw and that use of thawed samples does not introduce a significant bias for the measurement procedure. If use of frozen aliquots is not a viable means for sharing samples, then either an alternative sample transport approach is needed or else the study design and data analysis model must be adapted to use local sample collection at each site. Consult with a statistician and regulatory agencies as appropriate in such cases.

It is possible to perform the **multisite evaluation** by adopting at one site the 20 × 2 × 2 protocol (described in Chapter 3) or a suitable variant thereof, while adopting at two other sites a 5 × 5 (five days, five replicates per day) or 5 × 2 × 3 (five days, two runs per day, three replicates per run) protocol. This and other reasonable combinations yield unbalanced experimental designs and datasets, thus complicating the analysis. However, they can still be analyzed using ANOVA and comparable routines available in many statistical software applications. For such an approach, consult with a statistician as needed; the relevant data analysis is not treated here.

If use of **frozen aliquots** is not a viable means for sharing samples, then either an alternative sample transport approach is needed or else the study design and data analysis model must be adapted to use local sample collection at each site. Consult with a statistician and regulatory agencies as appropriate in such cases.

© Clinical and Laboratory Standards Institute. All rights reserved.

Matrix—Test samples for this protocol should be selected to reflect the characteristics of actual samples that would routinely be presented to a user's laboratory for testing. These include both patient specimens and control materials. It may be necessary to spike or dilute patient samples in order to achieve measurand levels in the high and low regions of the measuring interval where native patient samples cannot be sourced. Such contrived samples should be shown to be commutable with native patient samples. When necessary, stable, commercially available, protein-based materials may be used.

Measurand Levels—Samples representing a sufficient number of measurand levels must be tested in the study to adequately characterize the measurement procedure's precision over most of its stated measuring interval, including points at or near medically important decision levels within it. It is recommended that the samples prepared for use be screened at the central site before the start of the main experiment to check that they represent a suitable spectrum of values for evaluating the assay's precision performance. The number and positioning of these levels tested should be sufficient to ensure that one can extract from the tabulation reasonable precision estimates at intermediate levels, not just at measurand levels represented by the reported estimates.

In many cases, five or more levels will be needed to accommodate measurement procedures with wide measuring intervals and multiple medical decision points or medically important levels near the lower LoQ. Three levels may suffice for measurement procedures with limited measuring intervals and only one medical decision point or with consistent precision across all or most of the measuring interval. For assays intended for monitoring, measurand levels should be chosen carefully to reflect the intended use. For measurement procedures with medically important points at the assay's LoQ and/or LoD, it is recommended to include in the precision study measurand levels close to these clinically important points and to use numerical values below the LLMI in an estimation of precision (for LoQ) and percent of results "detected" (for LoD) (for details, consult CLSI document EP17[34]). For other measurement procedures, measurand levels for samples in the high- and low-end regions of the measuring interval should be selected such that most of the test results in the study would be expected to fall within the measuring interval of the measurement procedure. As such, it is not appropriate to use a sample with a measurand level too close to the ULMI or LLMI.

REMINDER:

See Section 3.4.5 on aliquots for guidance on whether or not all measurements for a sample in a given run may come from the **same aliquot.**

Aliquots—When appropriate and feasible, it is preferable to provide samples as aliquots of stable, frozen pools in order to minimize sample handling–related sources of variation within and across the different sites. Ensure that consistent practices for thawing and handling the aliquots are followed over the study duration at all participating sites. **NOTE:** See Section 3.4.5 on aliquots for guidance on whether or not all measurements for a sample in a given run may come from the same aliquot.

© Clinical and Laboratory Standards Institute. All rights reserved.

4.5 Experimental Steps

1. Calibrate the measurement procedure following the manufacturer's instructions at the start of the study. Periodically recalibrate during the study at the frequency designated by the manufacturer.

2. Perform a single run on "day one" of the study, generating five measurements for each sample. Include QC sample(s) within each run.

   NOTE: See Section 3.4.5 on aliquots for guidance on whether or not all measurements for a sample in a given run may come from the same aliquot.

3. Record the measurement obtained for each replicate, if possible using (at least) one more decimal place than would normally be used for reporting patient results, in order to minimize roundoff artifacts in the statistical analysis.

4. Repeat steps 2 and 3 for another four days of testing. The days do not have to be consecutive. The time of the runs does not have be the same each day. Indeed, if possible, the run times should be varied throughout the work day to introduce more randomness into the testing.

5. Review the results promptly each day to ensure that all tests were completed properly and that the measurement procedure continues to perform under acceptable statistical control. Assess the data for potential outliers, trending, error messages, or atypical results.

   NOTE 1: Any missing results for individual replicates—due, for example, to instrument malfunction or to unrectifiable operator or data recording errors—should be made up during the given run, if possible, and any missing runs should be repeated, preferably on the same day. If that is not possible, another full day should be added to the study for the site affected. When using the 3 × 5 × 5 design, the final dataset must consist of at least 25 measurements (representing five days or more) per sample per site—or 30 measurements when using the 3 × 5 × 2 × 3 design.

   NOTE 2: All values reported by the instrument are to be used unless specifically flagged with an error message that would preclude normal reporting, or if an error condition is identified that would explain an atypical result (eg, sample mishandling). Do **not** exclude any results from the final dataset without documented cause.

   NOTE 3: If a run must be rejected because of QC procedures (invalid run), conduct an additional run at the same day, if possible, after an investigation is conducted to identify and correct the

Any **missing results for individual replicates**—due, for example, to instrument malfunction or to unrectifiable operator or data recording errors—should be made up during the given run, if possible, and any missing runs should be repeated, preferably on the same day.

All values reported by the instrument are to be used unless specifically flagged with an error message that would preclude normal reporting, or if an error condition is identified that would explain an atypical result (eg, sample mishandling). **Do not exclude** any results from the final dataset without documented cause.

© Clinical and Laboratory Standards Institute. All rights reserved.

cause of the problem. Maintain a record of the number of invalid runs. If more than 5% of the runs need to be rejected and no assignable cause can be found, then the investigator should consider the possibility that the device is not sufficiently stable to allow for a valid assessment of its precision. It may be desirable to schedule additional days of evaluation at the outset of the investigation, to allow for potential run rejections, if needed.

4.6 Data Analysis

For ease of understanding, this section assumes that data analyses will be applied to balanced datasets. The text presents basic ANOVA calculations and interpretations. If the dataset for one or more of the test samples is unbalanced, it will be necessary to use a more advanced data analysis technique, as noted in Section 4.8.

4.6.1 Checking Data Integrity

Throughout the study, each of the participating sites should review their data periodically, preferably in real time, to ensure that measurements are being obtained as expected, the values appear reasonable, and the assumptions of the experiment are being met. Any suspect results should be identified as soon as possible to allow prompt follow-up investigation and, if warranted, repeat testing. Waiting until the end of the study to review the data may necessitate repeating the entire experiment for one or more samples if it is no longer possible to determine the assignable causes for highly discrepant measurements.

The final dataset for each sample should be reviewed once more after completion of the benchwork phase, by the central site. Histograms, trend plots, and other displays (see examples in Appendixes A and B) may be useful for inspecting the data at this stage, both on a site-by-site basis and in aggregate. Individual measurements or even clusters of measurements that seemed acceptable while the study was in progress might now stand revealed as potential outliers against the full body of results. They should be double-checked for transcription errors, sample mix-ups, or other anomalies. Carefully document all corrections and exclusions for assignable cause.

In the final analysis, to do justice to what the bulk of the data indicates about the assay's precision, it may be desirable to treat certain highly discordant measurements as "statistical outliers." This is acceptable providing that three conditions are met:

1. The discordancy of each result treated as a statistical outlier must be documented via an objective statistical outlier test.[27]

   The Grubbs' test is usually appropriate for the blocks (typically 5 × 5 = 25 or 5 × 2 × 3 = 30 results each) of site- and sample-specific data generated in a standardized reproducibility study. In CLSI document EP15,[1] it is the recommended technique for certifying outliers. By the Grubbs' test, in a dataset consisting of 25 (30)

 © Clinical and Laboratory Standards Institute. All rights reserved.

results, a result qualifies as an outlier at the 99% level if it lies more than 3.135 (respectively 3.236) SDs from the mean, when the SD is calculated from the data (including the suspect result) directly, ie, ignoring any structure induced by multiple sources of variation.[27,42]

NOTE: Even if a result qualifies in this manner as highly aberrant, it is not necessary to treat it as a statistical outlier. Doing so may not have any substantial impact on the precision estimates, and the fewer results treated as outliers the better.

Even if a result qualifies in this manner as **highly aberrant,** it is not necessary to treat it as a statistical outlier. Doing so may not have any substantial impact on the precision estimates, and the fewer results treated as outliers the better.

2. **A.** In any single block of data—consisting of the ≥ 25 results at a given site for a 3 × 5 × 5 experiment (or the ≥ 30 results at a given site from a 3 × 5 × 2 × 3 experiment)—no more than **one** result should be treated as a statistical outlier.

B. For the multisite, multisample study as a whole, there should be no more than **one** outlier per site if the number of samples tested in the study is ≤ 4—and no more than two outliers per site if the number of samples is > 4.

NOTE 1: Jointly, these two clauses allow for treating one result as a statistical outlier at any site for **any one** sample, but not for **every** site and sample in the study. In short, clause 2a applies locally, and clause 2b applies globally. For a multisample reproducibility study, the overall distribution of highly aberrant results is expected to be such that, at most, a few 25- or 30-result blocks will be affected, while many will not.

NOTE 2: Simplifying still further, this means that for a scenario involving 4 samples or less and N sites, at most N results may be treated as statistical outliers overall, and no two of these may come from the same site.

Any result treated as a statistical outlier belongs to the final dataset. Accordingly, data analysis for the precision study must be performed and summarized twice, ie, **both before and after** suppressing these discordant results, to demonstrate their impact on the precision estimates. The percentage of outliers should be reported; and their clinical significance should be evaluated.

Any result treated as a **statistical outlier** belongs to the final dataset. Accordingly, data analysis for the precision study must be performed and summarized twice, ie, both before and after suppressing these discordant results, to demonstrate their impact on the precision estimates.

4.6.2 Precision Estimates Calculated by Analysis of Variance

Estimates for variance components and the precision types of interest, along with CIs for the latter, are calculated one sample at a time, desirably with the help of existing software that is well documented and suitably vetted. (The sample-specific statistics should then be assessed in aggregate for what they have to say about the assay's precision characteristics across all or most of its measuring interval—minimally via a summary table, and desirably via precision profiles. See Section 4.7 and Appendix B, respectively.)

To allow for a) checking the consistency of repeatability and within-laboratory precision estimates across the sites sample by sample, and

© Clinical and Laboratory Standards Institute. All rights reserved.

b) obtaining estimates of reproducibility for each sample, two sets of calculations are required:

- Estimates of reproducibility based on the aggregate data are needed for every sample, and desirably CIs for these estimates as well:
  - When the 3 × 5 × 5 protocol is adopted, the reproducibility estimates and CIs can be obtained via two-way nested ANOVA implementations, hence by routines essentially the same as those featured in Sections 3.6.2 and 3.6.3 for the single-site 20 × 2 × 2 study, apart from differences in the n-sizes (study dimensions) and factor names involved. As a byproduct, the calculations yield for each sample what can be regarded as estimates of repeatability and within-laboratory precision "averaged" (in a certain way) across the participating sites: a valuable "frame of reference" for the site-specific estimates of these two precision types.
  - When the 3×5×2×3 protocol is adopted, the calculations can be handled by three-way nested ANOVA.[23]
- Site-specific repeatability and within-laboratory precision estimates for every sample also need to be determined from the final dataset for each participating site, as a basis for making site-to-site comparisons.
  - When the 3×5×5 protocol is adopted for the multisite study, for one way to handle these calculations, consult CLSI document EP15,[1] which spells out one-way ANOVA routines suitable for obtaining the site-specific repeatability and within-laboratory precision estimates—and CIs for the estimates—from 5×5 (five days, five results per day) datasets. Moreover, because the analyses in CLSI document EP15[1] accommodate unbalanced as well as balanced datasets for a spectrum of $n_{day} \times n_{rep}$ designs, they can deal with missing values and results suppressed as statistical outliers.
  - When the 3×5×2×3 protocol is adopted for the multisite study, two-way nested ANOVA can be applied.

In this section, discussion is limited to a sketch of the two-way nested ANOVA approach to the first set of calculations for complete—therefore balanced—datasets reflecting the 3×5×5 protocol (see the first bullet and sub-bullet, above). As noted elsewhere in this document, however, other suitable approaches to the data analysis are described in the literature, and some are implemented in reputable software applications. Moreover, many of the algorithms and implementations have the flexibility to deal with unbalanced datasets and with additional variants of the standardized designs featured in this document.

Table 3 displays a schematic ANOVA table representing, for any one sample, the pivotal output generated by most methods for fitting the two-factor, random-effects model formulated in Section 4.4 to a complete dataset (no missing values) arising from an experiment conforming to the 3 × 5 × 5 design. The specific values in the *DF* column are based on the recommended dimensions for this multisite design, namely, n_{site} = 3 sites, n_{day} = 5 days (one run per day), and n_{rep} = 5 replicates per run, implying a total of N = 75 results per sample.

© Clinical and Laboratory Standards Institute. All rights reserved.

Table 3. Generic Two-Way Nested ANOVA Summary Table Format, for Multisite 3 × 5 × 5

Source of Variation	*SS*	*DF*	*MS*	EMS
Site	SS_{site}	$DF_{site} = n_{site} - 1 = 2$	MS_{site}	$\sigma^2_{error} + n_{rep}\sigma^2_{day} + n_{day}n_{rep}\sigma^2_{site}$
Day (Site)	SS_{day}	$DF_{day} = (n_{day} - 1)n_{site} = 12$	MS_{run}	$\sigma^2_{error} + n_{rep}\sigma^2_{day}$
Error	SS_{error}	$DF_{error} = N - n_{site}n_{day} = 60$	MS_{error}	σ^2_{error}
Total	SS_{total}	$DF_{total} = N - 1 = 74$		

Abbreviations: *DF*, degrees of freedom; EMS, expected value of the mean square; MS, mean squares; SS, sum(s) of squares.

Some ANOVA routines will provide the associated model term variance components V_{site}, V_{day}, and V_{error}. If that information is not provided, calculate them from the ANOVA table values as follows.

$$V_{error} = MS_{error} \tag{14}$$

$$V_{day} = \text{the maximum of 0 or } \frac{(MS_{day} - MS_{error})}{n_{rep}} \tag{15}$$

$$V_{site} = \text{the maximum of 0 or } \frac{(MS_{site} - MS_{day})}{n_{day}n_{rep}} \tag{16}$$

As indicated in equations (15) and (16), "negative variance components" are dealt with in the ANOVA approach by resetting the estimates to zero. (In a REML analysis, they do not arise.)[23]

Three precision types are of interest for the multisite study, compared to just two for the single-site study. The repeatability SD (S_R) corresponds directly to the residual or error-term variance component (V_{error}), while the within-laboratory precision SD (S_{WL}) corresponds to the sum of the first two variance components, V_{error} and V_{day}, and the reproducibility SD (S_{REP}) corresponds to the sum of all three:

$$S_R = \sqrt{V_{error}} \tag{17}$$

$$S_{WL} = \sqrt{V_{day} + V_{error}} \tag{18}$$

$$S_{REP} = \sqrt{V_{site} + V_{day} + V_{error}} \tag{19}$$

These may be re-expressed as %CVs. Simply divide by the grand mean of all measurement results for the sample, then multiply by 100.

For a discussion of potentially applicable approaches to combining precision estimates, see Section 3.7.

REMINDER:

For a discussion of potentially applicable approaches to combining precision estimates, see Section 3.7.

© Clinical and Laboratory Standards Institute. All rights reserved.

4.6.3 Confidence Intervals for the Precision Estimates

The calculations in Sections 4.6.2 yield estimates for repeatability, within-laboratory precision, and reproducibility. From the estimates and their *DF*, we can construct CIs bracketing the estimates.

Assuming a balanced design and a complete dataset (no missing results), the *DF* for the repeatability estimate (S_R) can be calculated directly from the study's dimensions:

$$df_R = \mathrm{N} - n_{\mathrm{site}} n_{\mathrm{day}} \tag{20}$$

Approximating the *DF* for S_{WL} and S_{REP} requires more complex calculations. Due to the structure of the protocol, one cannot assume that all observations are independent. There are several different approaches to approximating df_{WL} and df_{REP}. Consult with a statistician to select a suitable method. The approach below uses Satterthwaite's original formula.[22]

$$df_{WL} = \frac{(\alpha_{\mathrm{day}} MS_{\mathrm{day}} + \alpha_{\mathrm{error}} MS_{\mathrm{error}})^2}{\frac{(\alpha_{\mathrm{day}} MS_{\mathrm{day}})^2}{DF_{\mathrm{day}}} + \frac{(\alpha_{\mathrm{error}} MS_{\mathrm{error}})^2}{DF_{\mathrm{error}}}} \tag{21}$$

$$df_{REP} = \frac{(\beta_{\mathrm{site}} MS_{\mathrm{site}} + \beta_{\mathrm{day}} MS_{\mathrm{day}} + \beta_{\mathrm{error}} MS_{\mathrm{error}})^2}{\frac{(\beta_{\mathrm{site}} MS_{\mathrm{site}})^2}{DF_{\mathrm{site}}} + \frac{(\beta_{\mathrm{day}} MS_{\mathrm{day}})^2}{DF_{\mathrm{day}}} + \frac{(\beta_{\mathrm{error}} MS_{\mathrm{error}})^2}{DF_{\mathrm{error}}}} \tag{22}$$

The *DF* terms come from the ANOVA data analysis report (see Table 3), as do the MS. The alpha and beta coefficients for this specific balanced protocol design are: $\alpha_{\mathrm{day}} = 0.5$, $\alpha_{\mathrm{error}} = 0.5$, $\beta_{\mathrm{site}} = 0.25$, $\beta_{\mathrm{day}} = 0.25$, and $\beta_{\mathrm{error}} = 0.50$. To determine the coefficient values appropriate for an unbalanced dataset, or if some other experimental protocol was adopted, consult a statistician or the references, or rely on suitably vetted software.

Illustrated below is the widely used chi-square (χ^2)–based approach to constructing CIs for precision estimates. (Other statistically valid approaches may also be used.) We are generally interested in 95% CIs, corresponding to an "alpha" level of 5% ($\alpha = 0.05$). In that case, at a CL of $1 - a = 0.95$, the lower and upper limits of a precision estimate's CI are calculated as:

$$s\sqrt{\frac{DF}{\chi^2_{(1-\frac{\alpha}{2}),DF}}} \quad \text{and} \quad s\sqrt{\frac{DF}{\chi^2_{\frac{\alpha}{2},DF}}} \tag{23}$$

where: $S =$ the precision estimate, expressed as an SD

$DF =$ the precision estimate's degrees of freedom

$\chi^2_{CL,DF} =$ chi-square distribution value for the desired CL and *DF*

Note that the pair of terms in equation (23) represents a two-sided interval. For the usual choice of $\alpha = 0.05$, this corresponds to a lower limit at $\chi^2_{0.975,DF}$ and an upper limit at $\chi^2_{0.025,DF}$. In some contexts, where only the upper limit is of interest, a "one-sided interval" may be more appropriate.

© Clinical and Laboratory Standards Institute. All rights reserved.

For a = 0.05, this corresponds to an upper confidence limit at $\chi^2_{0.05,DF}$.

Having constructed 95% CIs for the precision estimates expressed as SDs, simply divide these limits by the mean, and multiply by 100, to obtain **approximate** 95% CIs for the precision estimates expressed as %CVs. Providing the %CVs are less than 30% or 40%, this method of constructing CIs for them represents a satisfactory approach to characterizing their uncertainty as far as sampling variation is concerned. (See the two **NOTES** at the end of Section 3.6.3, which discuss the construction of CIs for %CVs.)

NOTE:

Reporting the results of a **multisite precision evaluation study in the assay's PI may be required** by the governing regulatory agency. If not, the study is reported in the PI per the manufacturer's discretion.

4.7 Suggested Format for Summarizing Multisite Precision Study Results

Reporting the results of a multisite precision evaluation study in the assay's PI may be required by the governing regulatory agency. If not, the study is reported in the PI per the manufacturer's discretion.

Similar to the report of a single-site study (see Section 3.8), the report of a multisite study should include a statement describing the evaluation protocol used. This should include all of the study's relevant elements, such as the number of sites, days, runs per day, and replicates per run; the total number of individual measurements; the number of missing values, invalid runs, invalid results, and results treated as statistical outliers.

A useful means to organize the basic statistics from the overall (across-site) analysis is through a summary table like the following.

Reproducibility														
Sample	Mean Value	N	Repeatability		Between-Run		Between-Day		Between-Site		Reproducibility	
			SD	%CV	SD	%CV	SD	%CV	SD	%CV	SD	%CV

Abbreviations: %CV, coefficient of variation expressed as a percentage; SD, standard deviation; N, total number of measurements obtained in the study (across sites, days, and replicates).

A summary of the within-site precision estimates can be accommodated in a separate table like the following, or in a combined table displaying both across-site and within-site statistics (see Table B3 in Appendix B for an example of within-site statistics).

© Clinical and Laboratory Standards Institute. All rights reserved.

Precision for Site 1										
Sample Description	Mean Value	N	Repeatability		Between-Run		Between-Day		Within-Laboratory	
			SD	%CV	SD	%CV	SD	%CV	SD	%CV

Abbreviations: %CV, coefficient of variation expressed as a percentage; SD, standard deviation; N, total number of measurements obtained in the study (across sites, days, and replicates).

Optional Information. Reporting some or all of the following information is optional, but often highly informative:

- *DF* for the repeatability, within-laboratory precision, and reproducibility estimates
- CIs for the repeatability, within-laboratory precision, and reproducibility estimates
- Precision profiles delineating the estimated relationship between precision and concentration for repeatability, within-run precision, and/or reproducibility (see examples in Appendix B)
- CBs for the precision profiles
- Diagnostic plots of the measurements by site and in order of generation (see, eg, Figures B1 and B2 in Appendix B)

4.8 Note on Unbalanced Datasets

As explained in Section 1.5.4.1 (in connection with Figure 3), the single-site 20 × 2 × 2 precision study and the multisite 3 × 5 × 5 reproducibility study represent fully balanced experimental designs, which can therefore be expected to yield for each sample a fully balanced dataset for analysis, providing that no results are missing or treated as statistical outliers. To keep the exposition simple, the discussion has made this assumption, and for the same reason, it has illustrated just one (widely adopted, conventional) approach to constructing CIs. For an unbalanced dataset—where, for example, not all runs have the same number of results or not all days have the same number of runs—a similar but somewhat different analysis applies. In particular, calculations differ for the EMS coefficients and for the *DF* displayed above in the ANOVA table and subsequent formulae.[31-33] Moreover, for constructing CIs, there are acceptable (arguably superior) alternatives to using the Satterthwaite/chi-square approach.[28-30,41] Software applications are available for generating the required precision estimates and CIs from unbalanced as well as balanced datasets.

© Clinical and Laboratory Standards Institute. All rights reserved.

Chapter 5
Conclusion

Conclusion

EP05-A3 provides both guidelines and two standardized protocols for the precision evaluation of quantitative measurement procedures used in clinical laboratories. This new edition retains the fundamental single-site protocol featured in earlier versions of the guideline; but it also introduces a multisite protocol, for reproducibility assessment. Moreover, it sketches additional methods (in Appendix C) for exploring an assay's precision characteristics—methods which can be implemented early in the development phase to gain a deeper grasp of the assay's performance. By including considerations on how to select and optimize the experimental protocol(s) best suited for a specific assay and its intended use, the document also seeks to impart to manufacturers and developers an understanding of the concepts and approaches involved.

© Clinical and Laboratory Standards Institute. All rights reserved.

Chapter 6

Supplemental Information

This chapter includes:

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

References

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

2. ISO. *Accuracy (trueness and precision) of measurement methods and results – Part 3: Intermediate measures of the precision of a standard measurement method*. ISO 5725-3. Geneva, Switzerland: International Organization for Standardization; 1994.

3. ISO. *Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*. ISO 5725-2. Geneva, Switzerland: International Organization for Standardization; 1994.

4. CLSI. *Evaluation of Stability of* In Vitro *Diagnostic Reagents; Approved Guideline*. CLSI document EP25-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2009.

5. CLSI. *Hemolysis, Icterus, and Lipemia/Turbidity Indices as Indicators of Interference in Clinical Laboratory Analysis; Approved Guideline*. CLSI document C56-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.

6. CLSI. *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline—Second Edition*. CLSI document EP12-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.

7. Miller JM, 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.

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

9. 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 12, 2014.

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

11. ISO. *In vitro diagnostic medical instruments – 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.

12. ISO. *In vitro diagnostic medical instruments – Measurement of quantities in biological samples – Metrological traceability of values for catalytic concentration of enzymes assigned calibrators and control materials*. ISO 18153. Geneva, Switzerland: International Organization for Standardization; 2003.

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

© Clinical and Laboratory Standards Institute. All rights reserved.

14 ISO. *Clinical laboratory medicine – In vitro diagnostic medical devices – Validation of user quality control procedures by the manufacturer.* ISO 15198. Geneva, Switzerland: International Organization for Standardization; 2004.

15 ISO. *Clinical laboratory testing and in vitro medical devices – Requirements for in vitro monitoring systems for self-testing of oral anticoagulant therapy*. ISO 17593. Geneva, Switzerland: International Organization for Standardization; 2007.

16 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.

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

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

19 Westgard JO. Statistical quality control procedures. *Clin Lab Med.* 2013;33(1):111–124.

20 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.

21 CLSI. *Expression of Measurement Uncertainty in Laboratory Medicine; Approved Guideline*. CLSI document EP29-A. Wayne, PA: Clinical and Laboratory Standards Institute, 2012.

22 Satterthwaite FE. Synthesis of variance. *Psychometrika.* 1941;6(5):309-316.

23 Searle SR, Casella G, McCulloch CE. *Variance Components*. New York, NY: John Wiley & Sons, Inc.; 2006.

24 Wilcox RR. *Introduction to Robust Estimation and Hypothesis Testing.* 3rd ed. San Diego, CA: Academic Press; 2012.

25 CLSI. *Preliminary Evaluation of Quantitative Clinical Laboratory Measurement Procedures; Approved Guideline—Third Edition.* CLSI document EP10-A3-AMD. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

26 ISO. *Accuracy (trueness and precision) of measurement methods and results - Part 5: Alternative methods for the determination of the precision of a standard measurement method*. ISO 5725-5. International Organization for Standardization; 1998.

27 Barnett V, Lewis T. *Outliers in Statistical Data*. 3rd ed. New York, NY: John Wiley & Sons, Inc.; 1994.

28 Chiang AKL. A simple general method for constructing confidence intervals for functions of variance components. *Technometrics*. 2001;43(3):356-367.

29 Lidong E, Hannig J, Iyer H. Fiducial intervals for variance components in an unbalanced two-component normal mixed linear model. *J Am Stat Assoc.* 2008;103(482):854-865.

30 General Satterthwaite approximation. In: Milliken GA, Johnson DE. *Analysis of Messy Data, Volume I: Designed Experiments.* 2nd ed. Boca Raton, FL: Taylor & Francis Group, LLC; 2009:348.

© Clinical and Laboratory Standards Institute. All rights reserved.

31 Burdick RK, Graybill FA. *Confidence intervals on variance components*. New York, NY: Marcel Dekker, Inc.; 1992.

32 Sahai H, Ojeda MM. *Analysis of Variance for Random Models, Volume I: Balanced Data – Theory, Methods, Applications, and Data Analysis*. New York, NY: Birkhäuser Boston; 2004.

33 Sahai H, Ojeda MM. *Analysis of Variance for Random Models, Volume II: Unbalanced Data – Theory, Methods, Applications, and Data Analysis*. New York, NY: Birkhäuser Boston; 2005.

34 CLSI. *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline—Second Edition*. CLSI document EP17-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.

35 Donner A, Zou GY. Closed-form confidence intervals for functions of the normal mean and standard deviation. *Stat Methods Med Res* 2012;21(4):347-359.

36 Distribution of the coefficient of variation. In: Hald A. *Statistical Theory with Engineering Applications*. New York, NY: John Wiley & Sons, Inc.; 1952:301-303.

37 Berman NG, Parker RA. Meta-analysis: neither quick nor easy. *BMC Med Res Methodol*. 2002;2:10.

38 Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. New York, NY: John Wiley & Sons, Inc.; 2009.

39 Sadler WA. Imprecision profiling. *Clin Biochem Rev*. 2008;29 Suppl 1:S33-36.

40 Sadler WA. Variance function program (version 11.0). http://www.aacb.asn.au/professionaldevelopment/useful-tools/variance-function-program-version-110. Accessed August 12, 2014.

41 Corbeil RR, Searle SR. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*. 1976;18(1):31-38.

42 Iglewicz B, Hoaglin DC. *How to Detect and Handle Outliers*. Milwaukee, WI: ASQC Quality Press; 1993.

© Clinical and Laboratory Standards Institute. All rights reserved.

Additional References

Ames MH, Webster JT. On estimating approximate degrees of freedom. *Am Stat.* 1991;45(1):45-50.

Gaylor DW, Hopper FN. Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics.* 1969;11(4):691-706.

Khuri AI. *Linear Model Methodology.* Boca Raton, FL: Taylor & Francis Group, LLC; 2010.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. *SAS for Mixed Models.* 2nd ed. Cary, NC: SAS Institute Inc.; 2006.

Liu JP, Lu LT, Liao CT. Statistical inference for the within-device precision of quantitative measurements in assay validation. *J Biopharm Stat.* 2009;19(5):763-778.

Mandel J. *The Statistical Analysis of Experimental Data.* New York. NY: Interscience Publishers; 1964.

Palma AZ. Verificación de la imprecisión empleando dos protocolos. *Rev Mex Patol Clin.* 2011;58(4):180-185.

Rao CR. Minimum variance quadratic unbiased estimation of variance components. *J Multivariate Anal.* 1971;1(4):445-456.

Robinson DL. Estimation and use of variance components. *J Roy Stat Soc D-Sta.* 1987;36(1):3-14.

Stöckl D, D'Hondt H, Thienpont LM. Method validation across the disciplines – critical investigation of major validation criteria and associated experimental protocols. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2009;877(23):2180-2190.

Vardeman SB, Jobe JM. *Statistical Quality Assurance Methods for Engineers.* New York, NY: John Wiley & Sons, Ltd.; 1998.

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix A. Worked Example—Single-Site Study

The dataset for this example consists of the results for just one sample in a single-site precision evaluation study, which in its entirety would involve the testing of several samples representing a broad spectrum of measurand levels. The dataset is taken from the previous editions of this document (EP05-A and EP05-A2) and is intended to provide a numerical example illustrating both the basic calculations and one graphical technique often useful in screening such data for anomalies.

The experimental design followed the classic 20 × 2 × 2 protocol for a single sample:

- One reagent lot
- One instrument system
- 20 test days
- Two runs per day
- Two replicate measurements per run

Measurements were obtained for all replicates (N = 80). Both replicates came from a single sample aliquot for each run, as there were no concerns with pretreatment or operator handling effects. There were no missing values. Accordingly, this is a **balanced** dataset, suitable for analysis via the standard analysis of variance (ANOVA) technique outlined both in the previous edition (EP05-A2) and in Section 3.6 of this document. The individual measurement results are listed in Table A1.

Table A1. Glucose Precision Evaluation Results (mg/dL)

		Run 1			Run 2	
Day		**Rep 1**	**Rep 2**		**Rep 1**	**Rep 2**
1		242	246		245	246
2		243	242		238	238
3		247	239		241	240
4		249	241		250	245
5		246	242		243	240
6		244	245		251	247
7		241	246		245	247
8		245	245		243	245
9		243	239		244	245
10		244	246		247	239
11		252	251		247	241
12		249	248		251	246
13		242	240		251	245
14		246	249		248	240
15		247	248		245	246
16		240	238		239	242
17		241	244		245	248
18		244	244		237	242
19		241	239		247	245
20		247	240		245	242

Abbreviation: Rep, replicate.

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix A. (Continued)

To screen for anomalies, the individual (single-replicate) results were plotted, concentration vs day, distinguishing by shape (and color) between the two runs performed each day. See Figure A1, which shows no apparent outliers, nor any evidence of trending in the data across days or time of day. (See also Appendix B, Section B4, which illustrates two minor variations on this graphical approach to screening for data integrity.)

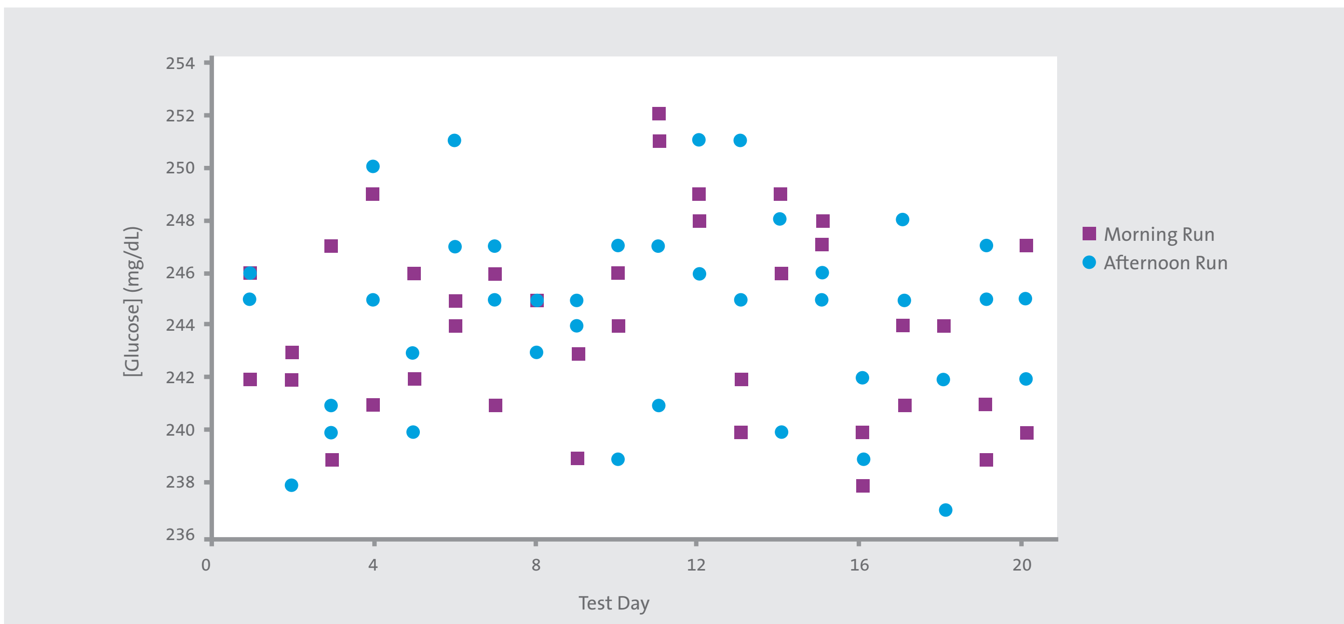


Figure A1. Glucose Precision Evaluation Results. Plots of the data from the morning runs (squares) and afternoon runs (circles) across 20 days for the glucose worked example. There are no points that stand out as atypical relative to the bulk of the data, nor is there any discernible trending (drift).

The data were analyzed in terms of the two-factor nested ANOVA model presented in Section 3.6.2 of this document, using a commercial software program for ANOVA. The program's principal output is summarized in Table A2.

Table A2. ANOVA Summary Table for the Glucose Example

Source of Variation	*SS*	*DF*	*MS*
Day	415.8	19	21.88
Run (Day)	281.0	20	14.05
Error	316.0	40	7.90
Total	1012.8	79	

Abbreviations: *DF*, degrees of freedom; *MS*, mean squares; *SS*, sum of squares.

Variance components are computed from the *MS* per equations (4) to (6) in Section 3.6.2 of this document, using $n_{rep} = 2$ and $n_{run} = 2$:

$$V_{error} = MS_{error} = 7.90 \quad \text{(A1)}$$

$$V_{run} = (MS_{run} - MS_{error}) / n_{rep} = (14.05 - 7.90) / 2 = 3.08 \quad \text{(A2)}$$

$$V_{day} = (MS_{day} - MS_{run}) / n_{run} n_{rep} = (21.88 - 14.05) / (2 \cdot 2) = 1.96 \quad \text{(A3)}$$

Because neither V_{run} nor V_{day} is less than zero, these values are taken without change.

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix A. (Continued)

The three variance component estimates are used to calculate estimates for the repeatability (S_R) and within-laboratory precision (S_{WL}) SDs per equations (7) and (8) in Section 3.6.2 of this document:

$$S_R = \sqrt{V_{error}} = \sqrt{7.90} = 2.81 \text{ mg/dL} \quad (A4)$$

$$S_{WL} = \sqrt{V_{day} + V_{run} + V_{error}} = \sqrt{1.96 + 3.08 + 7.90} = 3.60 \text{ mg/dL} \quad (A5)$$

Degrees of freedom (*DF*) for the repeatability SD are calculated from equation (9) in Section 3.6.3 of this document as:

$$df_R = \text{N} - n_{day} n_{run} = 80 - 20 \cdot 2 = 40 \quad (A6)$$

DF for the within-laboratory SD were calculated per equation (10) in Section 3.6.3 of this document as:

$$df_{WL} = \frac{(\alpha_{day} MS_{day} + \alpha_{run} MS_{run} + \alpha_{error} MS_{error})^2}{\frac{(\alpha_{day} MS_{day})^2}{DF_{day}} + \frac{(\alpha_{run} MS_{run})^2}{DF_{run}} + \frac{(\alpha_{error} MS_{error})^2}{DF_{error}}} \quad (A7)$$

$$df_{WL} = \frac{(0.25 \cdot 21.88 + 0.25 \cdot 14.05 + 0.5 \cdot 7.90)^2}{\frac{(0.25 \cdot 21.88)^2}{19} + \frac{(0.25 \cdot 14.05)^2}{20} + \frac{(0.5 \cdot 7.90)^2}{40}} = \frac{167.250}{1.575 + 0.617 + 0.390} \quad 64.8 \approx 65$$

The grand mean for all 80 test results is $\bar{x} = 244.2$ mg/dL. Using this value, the estimates expressed as SDs can be re-expressed as coefficients of variation expressed as percentages (%CVs):

$$\%\text{CV}_R = (S_R/\bar{x}) \cdot 100 = (2.81/244.2) \cdot 100 = 1.2\% \quad (A8)$$

$$\%\text{CV}_{WL} = (S_{WL}/\bar{x}) \cdot 100 = (3.60/244.2) \cdot 100 = 1.5\% \quad (A9)$$

The table below summarizes the minimally required statistics (see Section 3.8.1) in a format frequently used to report precision estimates in a package insert.

Sample Description	Mean (mg/dL)	Repeatability		Within-Laboratory Precision	
		SD	%CV	SD	%CV
Patient Pool	244.2	2.81	1.2%	3.60	1.5%

Abbreviations: %CV, coefficient of variation expressed as a percentage; SD, standard deviation.

Two-sided 95% confidence intervals (CIs) are computed for the repeatability and within-laboratory SD estimates per equation (11) in Section 3.6.3 of this document. Inputs are the precision estimates and their associated *DF*, from equations (A4) to (A7); the chi-square distribution values for the *DF* estimates; and the desired confidence level corresponding to α = 0.05. These are summarized in Table A3.

Table A3. Inputs for Calculating 95% CIs for the Glucose Example

		Repeatability		Within-Laboratory Precision
Precision Estimate (*S*)		2.81 mg/dL		3.60 mg/dL
DF		40		65
$\chi^2(1-\alpha/2), DF$		59.3		89.2
$\chi^2(\alpha/2), DF$		24.4		44.6

Abbreviation: *DF*, degrees of freedom.

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix A. (Continued)

Substituting these values into equation (11) in Section 3.6.3 of this document for repeatability gives:

$$2.81\sqrt{\frac{40}{59.3}} = 2.31 \text{ mg/dL and } 2.81\sqrt{\frac{40}{24.4}} = 3.60 \text{ mg/dL} \tag{A10}$$

Similarly, the lower and upper 95% confidence limits for within-laboratory precision are:

$$3.60\sqrt{\frac{65}{89.2}} = 3.07 \text{ mg/dL and } 3.60\sqrt{\frac{65}{44.6}} = 4.35 \text{ mg/dL} \tag{A11}$$

Analogous to the simple SD-to-%CV conversions illustrated in equations (A8) and (A9), the 95% confidence limits for the SD estimates can likewise be re-expressed as %CVs:

Repeatability:	0.95 to 1.5 %CV	(A12)
Within-Laboratory Precision:	1.3 to 1.8 %CV	

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix B. Worked Example—Multisite Study

B1 Overview

The dataset for this example (listed in its entirety in Table B1) represents a complete three-site, multisample reproducibility study of an automated immunometric assay intended for the quantitative measurement of CA19-9 in serum.

Samples—All three sites used the same samples. The samples were prepared, aliquotted, and frozen at Site 1 and distributed to the other two sites. They were chosen to be representative of the types of sample and the spectrum of concentrations encountered in routine practice. The samples, numbered in order of increasing concentration, included both human serum pools (P1, P2, P5) and control samples (Q3, Q4, Q6) of the kind recommended for use with the assay for internal QC monitoring. Jointly, as shown in Figures B3 and B4, the six samples represent four or five distinct measurand levels spanning a substantial fraction of the assay's stated 3 to 1000-kU/L measuring interval.

Benchwork—At each of the three sites, aliquots of the six samples were assayed as unknowns, in the same manner as patient samples, on each of five days, with one run per day, and five replicates per run (each replicate from a different aliquot), yielding 25 measurements per sample per site—for a total of 18 25-result blocks and 450 individual measurements. The study involved a different instrument and a different operator at each site, but all sites used the same reagent lot and the same calibrator lot. All runs included additional QC materials, serving to monitor the integrity of the individual runs. At each site, the assay was calibrated just once, shortly before the start of the first run, and not thereafter, because the assay had a recommended two-week calibration interval. Benchwork at the three sites was completed within essentially the same time period, in a total of eight calendar days.

Data Processing—On a day-to-day basis, each site monitored its own results for integrity, taking appropriate action. Sites 2 and 3 transmitted their results to Site 1 on completion of the benchwork. The combined three-site, six-sample dataset was then reviewed for integrity, analyzed and summarized at Site 1, using a spreadsheet facility in conjunction with programs for one- and two-way ANOVA and software dedicated to precision profile construction. The data processing was conducted in three stages:

- Global screen for highly aberrant points (potential outliers), drift, and other anomalies (see Section B2)
- Estimation for each sample, by ANOVA, of variance components and the precision types of interest, along with 95% confidence intervals (CIs) for the precision estimates (see Section B3)
- Visual examination, via precision profiles, of the sample-specific statistics in aggregate (see Section B4)

Because a comparison of site-to-site performance is of major interest for a reproducibility study, the most immediately informative summary is arguably Figure B4, which shows the **overall** within-laboratory precision estimates and their CIs (generated by two-way nested ANOVA from the combined dataset), together with a fitted curve and confidence band (CB), in relation to the site-specific within-laboratory precision estimates for each of the three participants (generated by one-way ANOVA from the **site-specific** data subsets). The former provides a sense of the "average" within-laboratory precision achievable across sites—on the assumption that the sites are representative, of course, and bearing in mind the relatively short duration of the study and the small number of participants involved—while (with the same stipulations) the three sets of site-specific estimates provide a sense of the dispersion of within-laboratory estimates that might arise, eg, when spot-checking manufacturer claims via five-day precision verification studies like those in CLSI document EP15.[1]

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix B. (Continued)

Table B1. CA19-9 Precision Evaluation Results (kU/L) for the Worked Example. Six samples, P1 to Q6, where "P" signifies patient sample pool and "Q" signifies QC material.

Day	Sample	Site 1	Site 2	Site 3	Sample	Site 1	Site 2	Site 3	Sample	Site 1	Site 2	Site 3
1	P1	12.5	12.8	12.2	Q3	59.0	53.9	58.4	P5	386.6	380.0	390.3
1	P1	11.8	12.8	11.7	Q3	57.7	55.1	57.9	P5	393.9	382.5	393.4
1	P1	11.6	12.7	11.6	Q3	58.1	54.9	57.2	P5	392.8	368.5	388.1
1	P1	11.6	13.2	11.9	Q3	56.3	53.7	56.4	P5	383.3	379.7	386.2
1	P1	11.7	11.1	11.6	Q3	55.6	51.3	56.1	P5	388.0	361.7	378.1
2	P1	12.2	11.4	12.0	Q3	56.5	52.7	56.4	P5	389.0	366.8	391.7
2	P1	11.8	10.7	11.7	Q3	56.3	54.4	57.4	P5	394.3	371.7	384.3
2	P1	11.6	11.0	11.9	Q3	58.4	49.8	57.9	P5	374.2	363.3	377.0
2	P1	12.2	13.8	12.1	Q3	57.3	51.4	57.5	P5	379.8	379.1	380.4
2	P1	11.5	11.4	11.7	Q3	56.5	49.0	55.9	P5	368.1	366.7	371.6
3	P1	12.0	14.7	12.0	Q3	57.7	55.1	57.2	P5	381.0	375.1	380.6
3	P1	11.9	13.7	11.7	Q3	56.7	55.0	57.7	P5	380.1	382.1	377.7
3	P1	11.4	13.2	11.5	Q3	58.6	54.1	57.4	P5	375.3	371.3	375.2
3	P1	11.6	13.2	11.8	Q3	56.1	52.5	56.3	P5	375.0	374.3	387.6
3	P1	11.9	13.1	11.7	Q3	55.2	54.5	56.5	P5	375.0	370.1	373.2
4	P1	11.9	13.6	11.5	Q3	56.5	55.2	57.5	P5	400.2	372.7	393.2
4	P1	11.1	13.5	11.8	Q3	58.4	56.1	56.9	P5	386.2	376.1	375.4
4	P1	12.4	12.8	11.9	Q3	55.4	55.2	55.6	P5	364.6	363.9	367.9
4	P1	11.3	12.3	12.3	Q3	55.7	55.8	56.3	P5	371.1	371.4	378.1
4	P1	11.5	12.0	12.6	Q3	57.7	54.9	56.1	P5	371.3	370.8	378.3
5	P1	13.3	14.1	12.0	Q3	56.9	54.5	56.6	P5	385.1	365.7	387.9
5	P1	10.7	12.8	10.5	Q3	56.3	55.9	56.3	P5	390.6	371.2	388.4
5	P1	10.3	12.2	10.5	Q3	56.3	51.7	56.0	P5	386.2	387.5	381.8
5	P1	10.7	15.8	11.2	Q3	55.7	50.0	55.7	P5	377.4	385.5	382.7
5	P1	11.9	13.3	11.1	Q3	55.0	54.2	57.0	P5	384.9	367.3	383.7
1	P2	45.0	40.6	45.1	Q4	175.0	156.6	171.2	Q6	433.9	410.9	434.3
1	P2	45.1	39.3	43.8	Q4	167.3	157.7	168.5	Q6	434.6	416.3	429.8
1	P2	42.4	41.4	42.4	Q4	169.7	155.8	168.8	Q6	424.9	391.8	425.1
1	P2	42.4	40.3	42.4	Q4	167.8	160.8	167.6	Q6	425.2	390.7	422.9
1	P2	42.5	37.3	42.1	Q4	170.5	153.6	170.5	Q6	424.9	388.4	417.8
2	P2	42.4	38.2	42.1	Q4	167.3	162.7	167.0	Q6	420.6	401.2	429.1
2	P2	41.7	37.5	42.4	Q4	166.7	157.1	169.4	Q6	437.5	409.0	425.6
2	P2	43.0	38.7	42.4	Q4	172.0	156.1	171.1	Q6	413.6	397.0	414.9
2	P2	41.8	43.0	43.5	Q4	170.1	149.5	168.8	Q6	416.2	395.7	419.6
2	P2	41.6	40.7	41.9	Q4	170.5	157.0	169.4	Q6	410.7	384.1	414.3
3	P2	45.1	42.1	43.9	Q4	167.4	159.9	168.1	Q6	423.0	401.4	417.9
3	P2	42.7	41.3	42.2	Q4	168.8	163.8	169.7	Q6	412.7	411.9	419.7
3	P2	41.6	40.5	41.3	Q4	170.5	160.9	173.5	Q6	426.7	392.2	431.0
3	P2	41.0	38.8	41.4	Q4	176.5	164.4	174.9	Q6	435.3	409.3	435.6
3	P2	42.2	39.9	42.0	Q4	168.2	156.3	167.8	Q6	417.9	391.9	422.6
4	P2	41.8	40.9	43.1	Q4	173.3	162.2	172.4	Q6	435.9	395.9	430.4
4	P2	44.4	40.7	43.6	Q4	171.4	164.6	167.8	Q6	424.9	415.7	415.8
4	P2	42.8	39.6	42.1	Q4	164.1	161.5	165.5	Q6	406.6	400.5	416.2
4	P2	41.4	40.1	41.3	Q4	166.9	159.2	165.9	Q6	425.8	409.4	421.9
4	P2	41.8	39.5	42.8	Q4	167.4	157.9	166.0	Q6	427.2	390.8	415.2
5	P2	41.2	40.8	41.1	Q4	164.8	158.2	166.0	Q6	418.0	410.5	416.9
5	P2	40.9	40.6	41.0	Q4	167.1	162.9	167.2	Q6	415.8	399.7	412.5
5	P2	41.1	41.4	40.7	Q4	167.3	162.4	166.6	Q6	409.2	387.9	410.2
5	P2	40.2	41.4	41.4	Q4	165.8	159.6	168.2	Q6	411.1	398.9	418.0
5	P2	40.9	37.3	43.9	Q4	165.8	160.6	168.8	Q6	423.2	382.5	415.2

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix B. (Continued)

B2 Screen for Anomalies—All Samples

To survey the dataset in aggregate for unexpectedly deviant results and other anomalies, the analyst first prepared a Levey-Jennings–like display, Figure B1, representing the dataset in its entirety. For closer visual inspection, the analyst also constructed a sample-specific display for each of the six samples, as illustrated for one of them (the lowest sample, P1) in Figure B2. (Note that Figure A1 in Appendix A illustrates yet a third way of looking at precision study results when screening for data integrity.)

Overall, and in the sample-specific displays, no individual measurements stand out as highly aberrant relative to the bulk of the data. In short, no results appear spurious or might reasonably be treated as statistical outliers. Moreover, none of the 25-result blocks of site- and sample-specific data exhibited any apparent drift capable of distorting the precision analysis.

B2.1 Outlier Criterion

As an objective, parametric basis for certifying results as statistical outliers, the protocol adopted for this study specified the Grubbs' test for single outliers at a 99% level, as suggested in Section 4.6.1 of this document, to be applied separately to the 25-result block of data for each sample at each site. By the Grubbs' test, in a set of 25 measurements, a result qualifies as an outlier only if it lies more than 3.135 SDs from the dataset's mean. On this basis—as illustrated for sample P1 in Figure B2—no results in the study qualify as outliers.

NOTES:

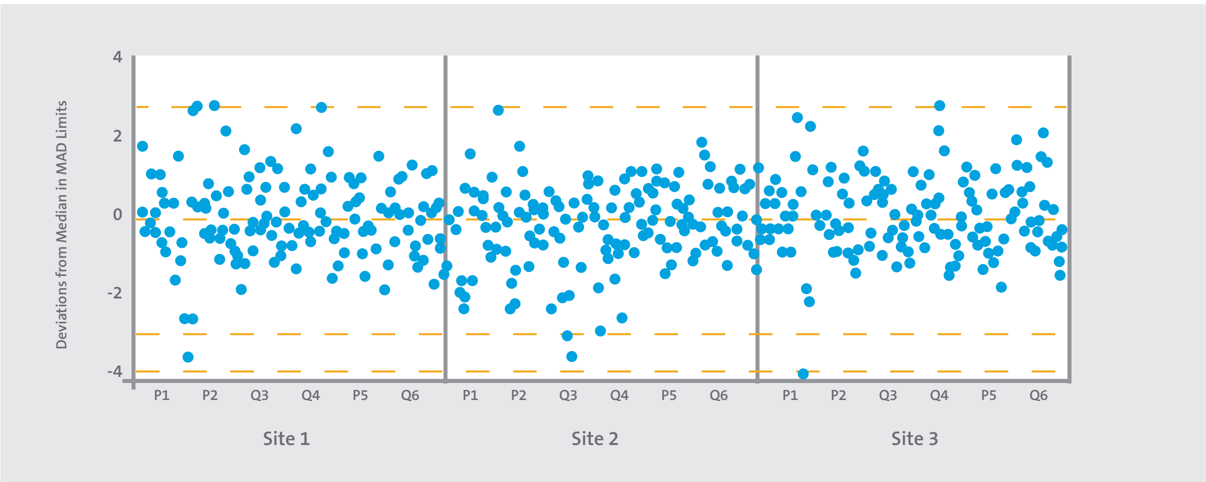
Figure B1: Along the horizontal axis, the individual results—3 × 6 × 25 = 450 in all—are plotted by site and sample, then in order of generation. The vertical axis is a Levey-Jennings scale, familiar from internal QC practice. There, it represents deviations from the mean in SD units. In the context of a multisample precision study, adopting such a scale allows for surveying all of the reproducibility study results in a single plot, even though they come from samples representing disparate measurand levels. Results expressed in terms of concentration are reduced to a Levey-Jennings scale by "standardizing" them. For a typical QC chart, this involves merely subtracting the sample's mean, then dividing by the sample's SD. Here, nonparametric counterparts of the mean and SD—namely, the median and the median absolute deviation (MAD), respectively[2]—were used for this step, on a sample-by-sample/within-site basis, specifically by subtracting the median for each 25-result block of data, then dividing by the MAD for that block. (Most spreadsheets have built-in functions for both median and MAD.) The vertical Levey-Jennings scale used for Figure B1 thus represents each measurement as a deviation, expressed in MAD units, from the median. The point of operating with robust proxies for the mean and SD in this context is to minimize the risk of masking, where a cluster of outliers, by inflating the parametric SD, can sometimes make itself invisible to common parametric outlier identification techniques, like the Grubbs' test.

Figure B2: The 3 × 25 = 75 individual results for a representative sample, namely P1, are plotted in native concentration units (kU/L), by site and then in order of generation. Superimposed on the data are site-specific horizontal lines providing two frames of reference for visually judging outlyingness:

- The **parametric** site- and sample-specific Grubbs' test limits described above, namely mean ± 3.135 SD, are indicated by the two outermost solid lines. The means are also delineated. This renders (when needed) outlier certification by the Grubbs' test a byproduct of this graphical exploratory investigation.
- Selected **nonparametric** gradations, corresponding to the robust Levey-Jennings scale in Figure B1 and useful for orientation, are indicated by dotted lines. Specifically, they indicate median ± 0, 3, and 4 MAD, where median and MAD can be thought of as robust proxies for the mean and SD (see the note on Figure B1 for details).

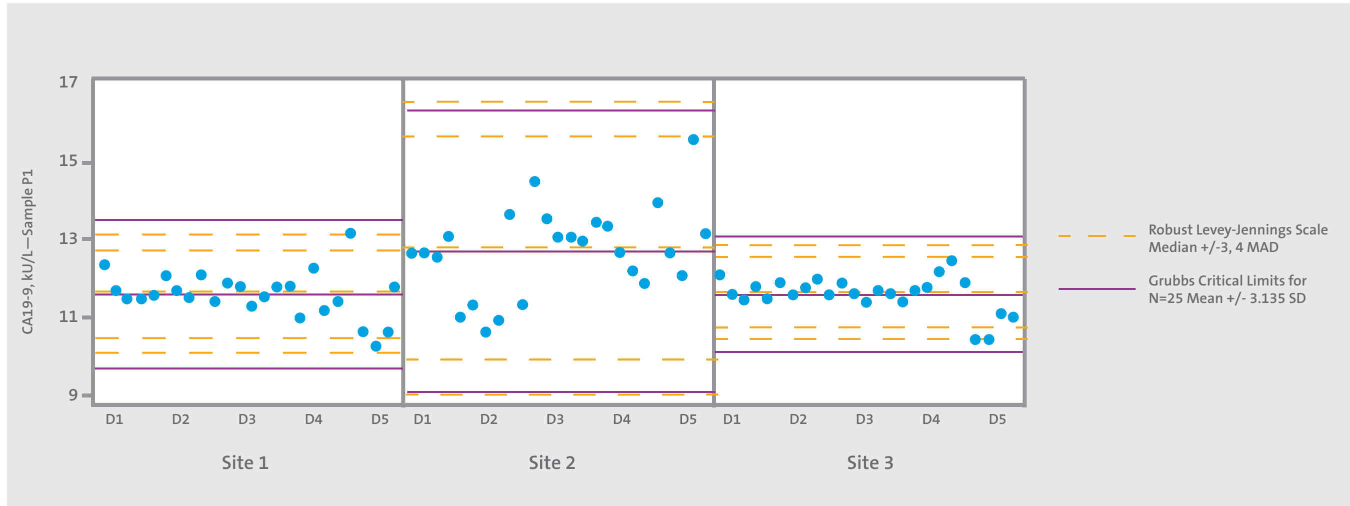
© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix B. (Continued)



Abbreviation: MAD, median absolute deviation.

Figure B1. Screen for Anomalies—All Samples. The vertical axis represents a robust Levey-Jennings scale.



Abbreviations: MAD, median absolute deviation; SD, standard deviation.

Figure B2. Screen for Anomalies—Sample P1 (Similar plots for the other five samples are not shown.)

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix B. (Continued)

B3 Analysis of Variance

With no missing results, and no results warranting treatment as statistical outliers, the dataset for each sample is complete, thus allowing for analysis of the combined dataset by the balanced two-way nested ANOVA calculations presented in Sections 4.6.2 and 4.6.3 of this document. The general purpose program adopted for this task was chosen in part, however, because of its ability to handle unbalanced as well as balanced datasets, should the need arise due to a missing result or statistical outlier. The three site-specific data subsets were analyzed by another program implementing the one-way random-effects ANOVA calculations described in CLSI document EP15.[1]

The analysis yielded estimates for the three variance components featured in Section 4.6.2 of this document —V_{error}, V_{day}, and V_{site}—corresponding respectively to the sources of variation identified in Table 3 of this document as "Error" (within-day or within-run), "Day (Site)" (between-day, for short, but more precisely: day-to-day/within-site), and "Site" (between-site). These estimates are listed in Table B2. Of potential interest to the manufacturer or developer, they suggest that the between-site component represents a substantial source of variation, whereas the between-day component is comparatively minor. (Bear in mind, however, that in the standardized three-site, five-day reproducibility study, only a small number of degrees of freedom [*DF*] are associated with these components.)

Of greater value to the clinical laboratory, the analysis also yielded estimates for the three precision types of interest (each a linear combination of the variance components), namely, repeatability, within-laboratory (within-site) precision, and reproducibility, calculated as described in Section 4.6.2 of this document. Expressed both as SDs—S_R, S_{WL}, and S_{REP}, respectively—and as coefficients of variation expressed as percentages (%CVs), these are tabulated in the topmost block of Table B3. The other three blocks provide comparable estimates for just two of the precision types, based on one-way ANOVA applied to the site-specific data subsets. Reproducibility cannot be estimated from the data for a single site, of course.

Output from the two-way analysis included 95% CIs for the overall precision estimates listed in the topmost block of Table B3 for the three precision types of interest. Calculated as described in Section 4.6.3 of this document, the CIs are listed in Table B4 for SDs and %CVs. They provide information on the uncertainty of the estimates as far as experimental size and "sampling variation" is concerned.

CIs generated by the one-way analysis for the site-specific repeatability and within-laboratory precision estimates are not tabulated in this appendix. The *DF* associated with any of the precision estimates are also not shown.

Table B2. Two-Way Nested ANOVA—Variance Components and Their Relative Magnitudes. For each component, V_i, the variances are tabulated in units of (kU/L) and as percent of total variance.[2]

Sample, kU/L		V_{error} (Within-day)		V_{day} (Between-day)		V_{site} (Between-site)	
P1	12.1	0.52	48.3%	0.18	16.4%	0.38	35.4%
P2	41.6	1.63	48.4%	0.12	3.6%	1.62	47.9%
Q3	55.7	1.56	29.7%	0.52	10.0%	3.17	60.4%
Q4	166	7.81	19.7%	1.87	4.7%	30.1	75.7%
P5	379	57.0	67.0%	3.19	3.7%	24.9	29.3%
Q6	414	74.0	30.7%	3.02	1.3%	164	68.1%

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix B. (Continued)

Table B3. Precision Estimates: Overall and Site by Site. The topmost block summarizes the two-way nested ANOVA analysis of the combined dataset. This is followed by a one-way ANOVA analysis for the data per site. Mean and SD units are kU/L.

		Repeatability		Within-Laboratory Precision		Reproducibility	
Sample	**Mean**	**SD**	**%CV**	**SD**	**%CV**	**SD**	**%CV**
P1	12.1	0.724	6.0%	0.838	6.9%	1.04	8.6%
P2	41.6	1.28	3.1%	1.33	3.2%	1.84	4.4%
Q3	55.7	1.25	2.2%	1.44	2.6%	2.29	4.1%
Q4	166	2.80	1.7%	3.11	1.9%	6.30	3.8%
P5	379	7.55	2.0%	7.76	2.0%	9.22	2.4%
Q6	414	8.60	2.1%	8.77	2.1%	15.5	3.7%
P1	11.7	0.647	5.5%	0.647	5.5%	Site 1	
P2	42.3	1.14	2.7%	1.39	3.3%		
Q3	56.8	1.15	2.0%	1.15	2.0%		
Q4	169	2.91	1.7%	3.09	1.8%		
P5	382	8.63	2.3%	9.06	2.4%		
Q6	422	8.46	2.0%	9.04	2.1%		
P1	12.8	1.01	7.9%	1.22	9.5%	Site 2	
P2	40.1	1.58	3.9%	1.58	3.9%		
Q3	53.6	1.67	3.1%	2.09	3.9%		
Q4	159	3.17	2.0%	3.68	2.3%		
P5	373	7.30	2.0%	7.30	2.0%		
Q6	399	10.6	2.7%	10.6	2.7%		
P1	11.7	0.374	3.2%	0.501	4.3%	Site 3	
P2	42.4	1.06	2.5%	1.10	2.6%		
Q3	56.8	0.748	1.3%	0.771	1.4%		
Q4	169	2.20	1.3%	2.43	1.4%		
P5	382	6.57	1.7%	7.01	1.8%		
Q6	421	6.21	1.5%	7.24	1.7%		

Abbreviations: %CV, coefficient of variation expressed as a percentage; SD, standard deviation.

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix B. (Continued)

Table B4. 95% CIs. These apply to the overall precision estimates listed in the topmost block of Table B3. SD units are kU/L.

	Repeatability		Within-Laboratory Precision		Reproducibility	
Sample	SD	%CV	SD	%CV	SD	%CV
P1	0.615–0.882	5.1% to 7.3%	0.703–1.04	5.8% to 8.6%	0.742–1.75	6.1% to 14.5%
P2	1.09–1.56	2.6% to 3.8%	1.14–1.59	2.7% to 3.8%	1.23–3.60	3.0% to 8.7%
Q3	1.06–1.52	1.9% to 2.7%	1.21–1.79	2.2% to 3.2%	1.43–5.70	2.6% to 10.2%
Q4	2.37–3.40	1.4% to 2.0%	2.63–3.81	1.6% to 2.3%	3.65–21.2	2.2% to 12.8%
P5	6.41–9.19	1.7% to 2.4%	6.65–9.30	1.8% to 2.5%	6.91–13.9	1.8% to 3.7%
Q6	7.30–10.5	1.8% to 2.5%	7.53–10.5	1.8% to 2.5%	9.35–43.7	2.3% to 10.6%

Abbreviations: %CV, coefficient of variation expressed as a percentage; SD, standard deviation.

B4 Precision Profiles

Precision profiles (see Figures B3 and B4) are provided as an aid in interpreting the tabulated results.

Precision profiles are graphs depicting the relationship between precision (strictly speaking: imprecision) and concentration (more generally; measurand level). Typically, the horizontal axis represents concentration, sometimes (as here) on a logarithmic scale, while the vertical axis represents precision expressed in terms of %CV (as here) or SD or variance. The term "precision profile" can also refer to a variance function fitted to the precision data, representing precision as a continuous function of concentration. The variance function can be displayed as a smooth curve in a plot of concentration vs precision.

The interest of such graphs lies, first, in the graphs' ability to summarize at a glance large amounts of tabular precision data. Thus, the points on display in Figure B3 represent the overall repeatability, within-laboratory precision, and reproducibility estimates tabulated in the top-most block of Table B3, while the points in Figure B4 summarize the site-specific within-laboratory precision estimates tabulated in the second, third, and fourth blocks of that table.

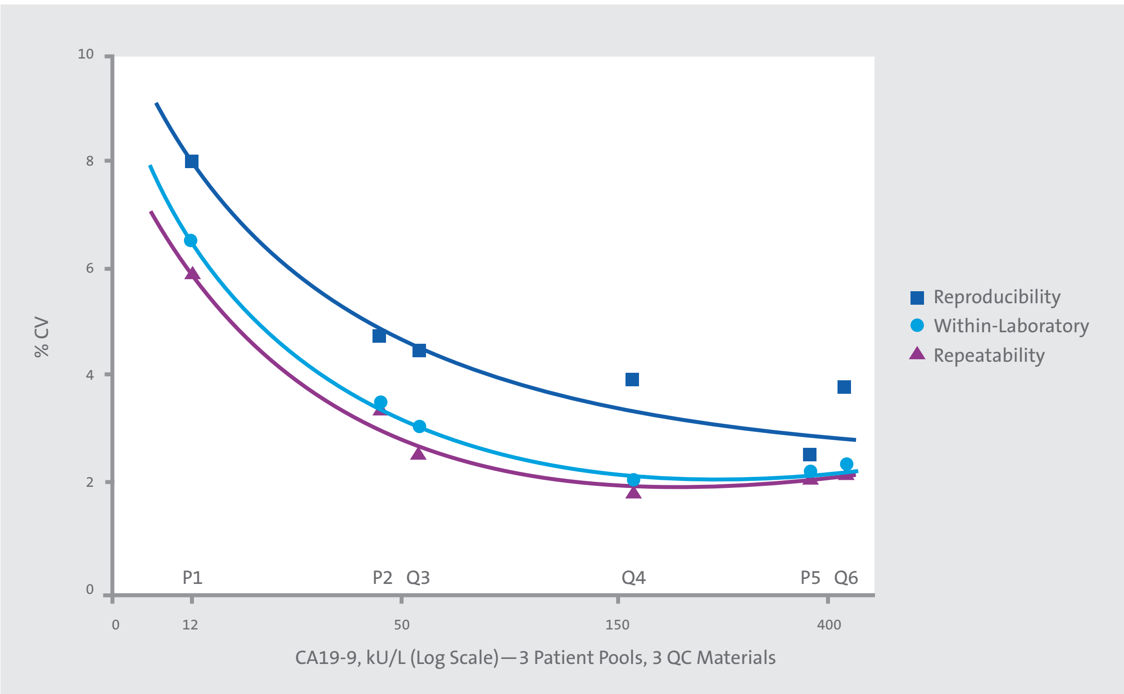
In addition, smooth curves (of an appropriate shape) fitted to the empirical data serve to model the presumptive underlying relationship between precision and concentration—analogous to the role of straight lines fitted to paired method comparison data. Here, curves of a shape suitable for immunoassay precision data have been fitted to the overall estimates (Table B3, topmost block) for each precision type in Figure B3, and to the overall estimates for within-laboratory precision, represented by crosses ("+" symbols), in Figure B4.

NOTES:

Note on Figure B3: Variance functions of the form Var = (a + b • X)—a three-parameter model suitable for typical immunoassays and immunometric assays[3,4]—were generated for the overall estimates of each precision type, using, as input to the program, the mean and SD values tabulated in the topmost block of Table B3. Note that the middle curve on display here is the same as the curve fitted to the crosses ("+" symbols) in Figure B4.

© Clinical and Laboratory Standards Institute. All rights reserved.

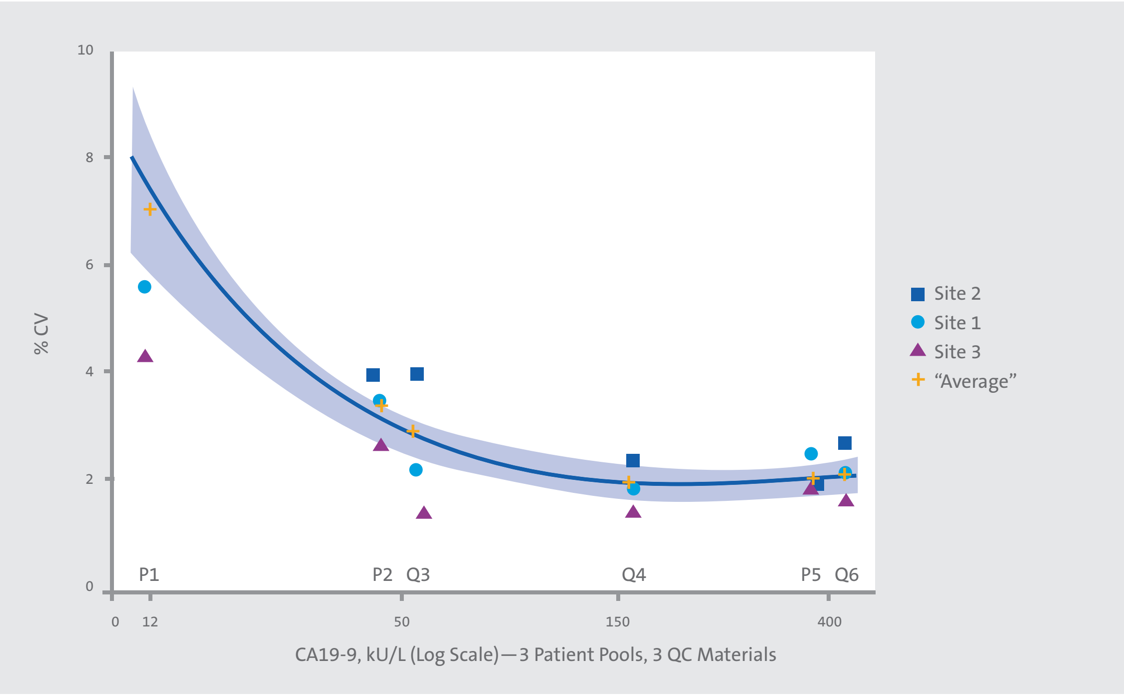
Appendix B. (Continued)



Abbreviations: %CV, coefficient of variation expressed as a percentage; QC, quality control.

Figure B3. Two-Way Nested ANOVA—Estimates

Note on Figure B4: The "+" symbols (crosses), used for plotting the overall within-laboratory precision estimates from the topmost block of Table B3, include vertical line segments ("needles") representing 95% CIs for those precision estimates. (Note that the curve fitted here to the crosses is the same as the middle curve on display in Figure B3.) The shaded area represents a 95% CB for the smooth curve (precision profile) fitted to the estimates. It provides a sense of the concentration-dependent uncertainty of the model for within-laboratory precision represented by the curve. And whereas the sample-specific CIs are independent of one another, the CB can be expected to shrink when the dataset underlying both the points on display and the fitted curve is amplified with additional means and SDs for samples exhibiting comparable precision performance.



Abbreviations: %CV, coefficient of variation expressed as a percentage; QC, quality control.

Figure B4. Within-Site (ie, Within-Laboratory) Precision Estimates Compared. Two-Way Nested ANOVA and One-Way ANOVA.

References for Appendix B

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

2. Wilcox RR. *Introduction to Robust Estimation and Hypothesis Testing.* 3rd ed. San Diego, CA: Academic Press; 2012.

3. Sadler WA. Imprecision profiling. *Clin Biochem Rev.* 2008;29 Suppl 1:S33-36.

4. Sadler WA. Variance function program. http://www.aacb.asn.au/professionaldevelopment/useful-tools/variance-function-program-version-110. Accessed August 12, 2014.

© Clinical and Laboratory Standards Institute. All rights reserved.

C1 Introduction

Table 1 in Section 2.2.2 of this document includes examples that cover major types of *in vitro* diagnostic (IVD) measurement procedures used in clinical laboratories. Based on these examples, the discussions in the following subsections will help developers of new measurement procedures to specify the model describing the data structure of measurement results, which is necessary for correctly extracting the relevant variance components and calculating the total variance estimate along with the associated degrees of freedom (*DF*).

The specific models described in the following sections are based upon a general model that includes the main effects of all the major analytical sources of variation on the measurement result $Y_{ijklmpqru}$ (obtained in *i*th site, on *j*th instrument, by *k*th operator, using *l*th reagent and *m*th calibrator lots for *p*th calibration on the *q*th day, in the *r*th run for the *u*th replicate), and assumes no significant interaction effects:

$$Y_{ijklmpqru} = \mu + S_i + I_{j(i)} + OP_{k(i)} + RL_l + CL_m + CC_{p(ijklm)} + D_{q(ijklmp)} + R_{r(ijklmpq)} + \varepsilon_{u(ijklmpqr)} \quad \text{(C1)}$$

The indices in parentheses denote inherent nesting of the respective effects. Note that the general model above is not fully nested; some effects are crossed—in particular, S, RL and CL. Correct specification of the nested and crossed effects in the model is important for correct estimation of the variance components along with the number of degrees of freedom of the total variance. Notation used in equation (C1) for the respective sources of variation is described in Table C1.

Table C1. Notation for Sources of Variation

Notation	Shift Due to	Causes for Variability	Nesting
S_i	ith ($1 \ldots n_S$) testing site	Site-specific sample preparation procedures, ambience, etc.	None
$I_{j(i)}$	jth ($1 \ldots n_I$) instrument	Variability in instrument effect not removable by calibration	Within site (shown), and optionally within operator
$OP_{k(i)}$	kth ($1 \ldots n_{OP}$) operator	Variability in manual sample preparation, etc.	Within site
RL_l	lth ($1 \ldots n_{RL}$) reagent lot	Degradation after calibration; sensitivity variability not removable by calibration due to lack of calibrator commutability with clinical specimens	Optionally within site
CL_m	mth ($1 \ldots n_{CL}$) calibrator lot	Value assignment errors and on-shelf degradation	Optionally within site, etc.
$CC_{p(ijklm)}$	pth ($1 \ldots n_{CC}$) calibration cycle	Random error of each individual calibration	Within all combinations of all the above sources
$D_{q(ijklmp)}$	qth ($1 \ldots n_D$) day	Variability of ambient conditions, sample degradation, etc.	Within all combinations of all the above sources
$R_{r(ijklmpq)}$	rth ($1 \ldots n_R$) run	Variability of ambient conditions, sample degradation, etc.	Within all combinations of all the above sources
$\varepsilon_{u(ijklmpqr)}$	random error of uth ($1 \ldots n_u$) replicates	Sources not included in the ANOVA model such as sample and reagent volume variability among tests, etc.	Within all combinations of all the above sources

Abbreviation: ANOVA, analysis of variance.

The model to be used for a particular study can be derived from the general model in equation (C1) by retaining only the sources of variation relevant to the study. For example, a model for one site, one instrument, one operator, one lot of reagent, one calibrator lot, and a single calibration (a basic study with day, run, and random sources of variation) retains

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix C. (Continued)

only the terms corresponding to these sources of variation, which, after re-ordering the indices for the effects, is the same as equation (1) in Section 1.5.4.1 of this document:

$$Y_{ijk} = \mu + D_i + R_{j(i)} + \varepsilon_{k(ij)}$$

In a multisite study, the site effect is inherently not nested in other effects. Instrument and operator effects are inherently nested in the site effect because the particular instruments and operators used are located at a particular site. Similarly, individual instrument calibrations are inherently nested in instruments. Day, run, and random effects are inherently nested in combinations of the effects of all other sources of variation preceding these in the general model described with equation (C1).

Individual variance components can only be estimated for the sources of variation that are not confounded with other sources of variation in the experimental design. As the number of variance components to be estimated increases, so does the size of the experiment. Therefore, understanding, based on prior experience or on scientific considerations, which of the sources of variation are expected to be major contributors to the total variability is important for designing an economical experiment that allows for estimation of the desired variance components. In many cases, more economical experiments can be designed by deliberately allowing confounding of minor sources of variation. For example, a single operator can run each instrument over the study, confounding the effects of the operators and instruments, when the measurement procedure is fully automated and no operator effect is expected.

The traditional model, used in all previous editions of this guideline and described in Chapter 3, included only day, run, and random components—even when several instruments, several lots of reagent and calibrator, several operators, and several calibrations per instrument might have been used in the study. These additional sources of variation would have been confounded with the day and run effects and not extracted separately. Such treatment of the study data helps to reduce the study size and simplify the statistical analysis, but may lead to biased estimates of variance components and incorrect DF calculations.

Day and run effects were considered to be the major sources of variation in early history of IVD tests. Reagents at that time typically had rather short stability and drifted over time, requiring frequent calibrations—often with each run. For most modern measurement procedures, however, reagents and calibrators have 12 to 18 months of shelf stability and the day and run sources of variation are no longer major contributors. While random variation (repeatability) remains a major contributor to total variance for modern automated assays, the manufacturing variability between lots of reagent and calibrator value assignments also may be major sources of variation to consider.

It is advantageous for developers to understand the relative contributions of different sources of variation to the total variability of a measurement procedure, both for establishing suitable control systems and to prioritize improvement projects to reduce the total variability. The size of experiments to allow quantifying such information, however, can become too large and/or not practical to attempt. Fortunately, some of these sources of variation may be known *a priori* to be negligible based upon technical knowledge of the measurement procedure. Therefore, while all the major sources of variation should be included in an evaluation of total precision, only a few specific sources may need to be quantified separately. Such an approach allows for realistic estimates of total variance with an experiment of reasonable size.

Various combinations of the numbers of sites, instruments per site, reagent lots, days per combination of site, instrument, reagent lot, and runs per day can be examined to select the most desirable experiment design. One criterion for a good design is having reasonable balance in the *DF* for the sources of variation to be quantified. As such, it may be better to use more reagent and/or calibrator lots and conduct the study over a shorter time period, as shown in the following two examples.

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix C. (Continued)

Example Design 1

Consider a precision study based upon the following design factors:

- n_{site} = two sites
- n_{inst} = two instruments per site
- n_{reag} = two reagent lots
- n_{day} = 20 days per site-instrument-reagent lot combination
- n_{run} = two runs a day for each site-instrument-reagent lot combination
- n_{rep} = two replicates in each run

In this design, as usual, days are nested within sites, runs are nested within days, and replicates are nested within runs. However, reagent lots are crossed with sites because both reagent lots are evaluated at both sites. The total number of tests per sample is given by:

n = 2 sites • 2 instruments • 2 reagent lots • 20 days • 2 runs/day • 2 replicates/run = 640 tests

The associated *DF* for each design factor are calculated to be:

Site $DF = n_{site} - 1 = 2 - 1 = 1$

Instrument $DF = n_{site}(n_{inst} - 1) = 2(2 - 1) = 2$

Reagent lot $DF = n_{reag} - 1 = 2 - 1 = 1$

Run $DF = n_{site}\, n_{inst}\, n_{reag}\, n_{day}(n_{run} - 1) = 2 \bullet 2 \bullet 2 \bullet 20(2 - 1) = 160$

Repeatability $DF = n_{site}\, n_{inst}\, n_{reag}\, n_{day}\, n_{run}(n_{rep} - 1) = 2 \bullet 2 \bullet 2 \bullet 20 \bullet 2(2 - 1) = 320$

Day $DF = n - DF_{site} - DF_{inst} - DF_{reag} - DF_{run} - (DF_{rep} - 1)$

$= 640 - 1 - 2 - 1 - 160 - 320 - 1 = 155$

Example Design 2

Now, consider a modified design for this precision study where the number of sites and reagent lots is increased but the number of days is decreased:

- n_{site} = three sites
- n_{inst} = two instruments per site
- n_{reag} = three reagent lots
- n_{day} = five days per site-instrument-reagent lot combination
- n_{run} = two runs a day for each site-instrument-reagent lot combination
- n_{rep} = two replicates in each run

The total number of tests per sample is given by:

n = 3 sites • 2 instruments • 3 reagent lots • 5 days • 2 runs/day • 2 replicates/run = 360 tests

The associated *DF* for each design factor are calculated to be:

Site $DF = n_{site} - 1 = 3 - 1 = 2$

Instrument $DF = n_{site}(n_{inst} - 1) = 3(2 - 1) = 3$

Reagent lot $DF = n_{reag} - 1 = 3 - 1 = 2$

Run $DF = n_{site}\, n_{inst}\, n_{reag}\, n_{day}(n_{run} - 1) = 3 \bullet 2 \bullet 3 \bullet 5(2 - 1) = 90$

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix C. (Continued)

$$\text{Repeatability } DF = n_{site}\, n_{inst}\, n_{reag}\, n_{day}\, n_{run}(n_{rep} - 1) = 3 \bullet 2 \bullet 3 \bullet 5 \bullet 2(2 - 1) = 180$$

$$\text{Day } DF = n - DF_{site} - DF_{inst} - DF_{reag} - DF_{run} - (DF_{rep} - 1)$$

$$= 360 - 2 - 3 - 2 - 90 - 180 - 1 = 82$$

While numbers of *DF* for day, run, and repeatability variance components are still large, ranging from 82 to 180, the numbers of *DF* for "site" and "lot" have doubled, and the number of *DF* for "instrument" has increased 1.5-fold, the total number of tests was reduced from 640 to 360 per sample—a savings of 44% in the amount of testing required per sample.

The models described in the following sections include relevant sources of variation as crossed effects and/or nested effects. Particular studies can use subsets of the sources of variation shown in the models. For example, in-house precision evaluations are often performed at a single site, allowing "site" to be excluded from the model. In practice, studies often are conducted with single lots each for reagent and calibrator, and a single calibration of each instrument used in the study. In such cases, the reagent lot, calibrator lot, and instrument calibration effects on random variation are not included in the study. In multisite studies, single but different lots of reagent or calibrator can be used in different sites. In such situations, the reagent and/or calibrator lot effects are present in test results but confounded with the site effect.

C2 Automated Device Calibrated Periodically

Periodic calibration of automated devices is a very common practice and includes most current general clinical chemistry and immunoassay measurement procedures (see Example 1 in Table 1 of Section 2.2.2 of this document). The typical time period between calibrations often falls between 14 days to 60 days, with infrequent irregularities caused by a shift in the QC chart or switching to a new lot of reagent.

Various combinations of the sources of variation in the general model can be included in this type of study. The most relevant sources are calibrator lot and instrument calibration. Day and run components are present because the studies are conducted over time. The calibrator lot effect consists of error due to value assignment and error due to degradation over time. Use of several lots of calibrator, preferably of different ages relative to shelf life, is desirable.

If the duration of the calibration cycle substantially exceeds that of the study, either more frequent calibrations can be done per instrument, or just a single calibration per instrument used. In the latter case, the effects of the instrument calibration and instrument will be confounded. Site is a factor in case of a multisite study. For certain types of assays, it is arguable that reagent lot should not be a large contributor to total variability, because its impact should be mitigated by calibration if the assay is well designed. For most automated devices operator effect is negligible. These considerations lead to the following model, with the relevant effects retained from the general model:

$$Y_{ijklmpq} = \mu + S_i + I_{j(i)} + CL_k + CC_{l(ijk)} + D_{m(ijkl)} + R_{p(ijklm)} + \varepsilon_{q(ijklmp)} \quad \text{(C2)}$$

The notation above is the same as in equation (C1), though reagent lot and operator were deleted and those for the other sources of variation adjusted. The experiment should be designed to allow for separate estimation of the variance components included in model.

C3 Automated Device Calibrated Frequently

Frequent calibration of automated devices is typical for assays (using ion-selective electrodes) of electrolytes, such as Na, K, Cl, and CO_2. The model shown in equation (C2) for the automated device calibrated periodically also applies to this case, though calibrations of each instrument are made following the manufacturer's manual daily or for each run.

© Clinical and Laboratory Standards Institute. All rights reserved.

Appendix C. (Continued)

C4 Automated Device That Is Not Calibrated

No calibration is typically for a small subset of assays that lack reliably value-assigned and stable calibrators. Examples of such assays are amylase, creatine kinase, and lactate dehydrogenase. A single fixed calibration factor is used for all reagent lots, calibrator lots, and instruments. Because the instrument is automated, there should be no significant operator effect on the measurement procedure. Operator-associated variability may, however, enter the experiment from sample handling variability and would be confounded with other sources of variation. The general model for this case can be written as:

$$Y_{ijklmp} = \mu + S_i + I_{j(i)} + RL_k + D_{l(ijk)} + R_{m(ijkl)} + \varepsilon_{p(ijklm)} \quad \text{(C3)}$$

C5 Automated Device With Reagent Lot–Specific Fixed Calibration

For automated devices with reagent lot–specific fixed calibration, calibration is fixed for each lot of reagent for all sites and instruments where it is used. The model in this case is the same as in the previous case (Section C4), but the reagent lot effect on total variability is reduced by using lot-specific calibration parameters.

C6 Automated Device With Reagent Lot–Specific Fixed Calibration and Pretreatment

In addition to the sources of variation important for the previous case (Section C5), operator effect in automated devices with reagent lot–specific fixed calibration and pretreatment may be significant, and is included in the model:

$$Y_{ijklmpq} = \mu + S_i + I_{j(i)} + OP_{k(i)} + RL_l + D_{m(ijkl)} + R_{p(ijklm)} + \varepsilon_{q(ijklmp)} \quad \text{(C4)}$$

C7 Manual or Semiautomated Device

The model for a manual or semiautomated device can be the same as described for the previous device in Section C6.

C8 Multiple Automated Devices Within a Single Laboratory

The model for multiple automated devices within a single laboratory is the same as the model in Section C4 but with no site effect included.

© Clinical and Laboratory Standards Institute. All rights reserved.

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

EP05-A3 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
						X					
						C24					
						C56					
						EP10					
						EP12					
						EP15					
						EP17					
						EP25					
						EP29					
		M29									

© Clinical and Laboratory Standards Institute. All rights reserved.

Related CLSI Reference Materials*

C24-A3 **Risk Management Techniques to Identify and Control Laboratory Error Sources; Approved Guideline—Second Edition (2009).** This guideline provides definitions of analytical intervals, planning of quality control procedures, and guidance for quality control applications.

C56-A **Hemolysis, Icterus, and Lipemia/Turbidity Indices as Indicators of Interference in Clinical Laboratory Analysis; Approved Guideline (2012).** This document provides background information on mechanisms of hemolysis, icterus, lipemia/turbidity (HIL) interference; intended usefulness of HIL indices; establishment of HIL alert indices; availability of automated HIL detection systems; and interpretation, strengths, limitations, and verification of HIL indices in the clinical laboratory.

EP10-A3-AMD **Preliminary Evaluation of Quantitative Clinical Laboratory Measurement Procedures; Approved Guideline—Third Edition (2014).** This guideline provides experimental design and data analysis for preliminary evaluation of the performance of a measurement procedure or device.

EP12-A2 **User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline—Second Edition (2008).** This document provides a consistent approach for protocol design and data analysis when evaluating qualitative diagnostic tests. Guidance is provided for both precision and method-comparison studies.

EP15-A3 **User Verification of Precision and Estimation of Bias; Approved Guideline—Third Edition (2014).** This document describes the estimation of imprecision and of bias for clinical laboratory quantitative measurement procedures using a protocol that can be completed within as few as five days.

EP17-A2 **Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline—Second Edition (2012).** This document provides guidance for evaluation and documentation of the detection capability of clinical laboratory measurement procedures (ie, limits of blank, detection, and quantitation), for verification of manufacturers' detection capability claims, and for the proper use and interpretation of different detection capability estimates.

EP25-A **Evaluation of Stability of *In Vitro* Diagnostic Reagents; Approved Guideline (2009).** This document provides guidance for establishing shelf-life and in-use stability claims for *in vitro* diagnostic reagents such as reagent kits, calibrators, and control products.

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

© Clinical and Laboratory Standards Institute. All rights reserved.

Related CLSI Reference Materials (Continued)

EP29-A **Expression of Measurement Uncertainty in Laboratory Medicine; Approved Guideline (2012).** This guideline describes a practical approach to assist clinical laboratories in developing and calculating useful estimates of measurement uncertainty, and illustrates their application in maintaining and improving the quality of measured values used in patient care.

M29-A4 **Protection of Laboratory Workers From Occupationally Acquired Infections; Approved Guideline—Fourth Edition (2014).** Based on US regulations, this document provides guidance on the risk of transmission of infectious agents by aerosols, droplets, blood, and body substances in a laboratory setting; specific precautions for preventing the laboratory transmission of microbial infection from laboratory instruments and materials; and recommendations for the management of exposure to infectious agents.

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

Industry and Large Commercial Laboratories

Abbott (IL)
Abbott Point of Care Inc. (NJ)
AdvaMed (DC)
Aria Diagnostics (CA)
ARUP Laboratories (UT)
Astellas Pharma (IL)
AstraZeneca Pharmaceuticals (MA)
Astute Medical, Inc. (CA)
Axis-Shield PoC AS (United Kingdom [GB])
Bayer Healthcare, LLC Diagnostic Division (IN)
BD (NJ)
Beckman Coulter, Inc. (PA)
Bioanalyse, Ltd. (Turkey)
Biohit Oyj. (Finland)
BioMerieux, Inc. (MO)
Bio-Rad Laboratories, Inc. (CA)
Canon U.S. Life Sciences, Inc. (MD)
Cempra Pharmaceuticals, Inc. (NC)
Cepheid (CA)
Abbott (IL)
Abbott Point of Care Inc. (NJ)
Accelerate Diagnostics Inc. (AZ)
AdvaMed (DC)
ARH Regional Medical Center (KY)
ARUP Laboratories (UT)
Astellas Pharma (IL)
AstraZeneca Pharmaceuticals (MA)
Astute Medical, Inc. (CA)
Axis-Shield PoC AS (United Kingdom [GB])
Bayer Healthcare, LLC Diagnostic Division (KS)
BD (NJ)
Beckman Coulter (PA)
Bio-Rad Laboratories, Inc. (CA)
Bioanalyse, Ltd. (Turkey)
Biohit Oyj. (Finland)
Biomedia (Thailand) Co.,Ltd. (Thailand)
BioMerieux, Inc. (MO)
Canon U.S. Life Sciences, Inc. (MD)
Cempra Pharmaceuticals, Inc. (NC)
Cepheid (CA)
Cerexa, Inc. (CA)
Clinical Reference Laboratory (MO)
Cubist Pharmaceuticals, Inc. (MA)
Eiken Chemical Company, Ltd. (Japan)
Elanco Animal Health (IN)
EMH Regional Medical Center (OH)
Enzo Clinical Labs (NY)
Exosome Diagnostics, Inc. (MN)
Greiner Bio-One GmbH (Austria)
Greiner Bio-One Inc. (NC)
Guangzhou Daan Clinical Laboratory Center Co. Ltd (China)
Himedia Labs Ltd (India)
Hinsdale Pathology Associates (IL)
Hologic, Inc. (MA)
Icon Laboratories, Inc. (NY)
Instrumentation Laboratory (MA)
Johnson & Johnson Pharmaceutical Research & Development, L.L.C. (NJ)
Kaiser Permanente (CA)
Laboratory Corporation of America (VA)
Life Laboratories (MA)
LifeLabs (Canada)
LifeLabs Medical Laboratory Services (Canada)
Mbio Diagnostics, Inc. (CO)
Melinta Therapeutics, Inc. (CT)
Merck & Company, Inc. (NJ)
Merial Limited & Newport Laboratories (MO)
Microbiologics (MN)
Micromyx, LLC (MI)
Myraqa, Inc. (CA)
Nihon Kohden Corporation (Japan)
Nissui Pharmaceutical Co., Ltd. (Japan)
Nova Biomedical Corporation (MA)
NovaBiotics (United Kingdom [GB])
Novartis Institutes for Biomedical Research (CA)
Ortho-Clinical Diagnostics, Inc. (NY)
Oxyrase, Inc. (OH)
PathCare Pathology Laboratory (South Africa)
PerkinElmer (Finland)
PerkinElmer Genetics, Inc. (PA)
Pfizer Inc (PA)
Phadia AB (Sweden)
Philips Healthcare Incubator (Netherlands)
QML Pathology (Australia)
Quest Diagnostics Nichols Institute (CA)
Radiometer Medical A/S (Denmark)
Roche Diagnostics Corporation (IN)
Sanofi Pasteur (PA)
Sarstedt, Inc. (NC)
Sekisui Diagnostics (MA)
Siemens Healthcare Diagnostics, Inc. (GA)
Sonic Healthcare USA (TX)
Streck Laboratories, Inc. (NE)
Sysmex America, Inc. (IL)
The Binding Site Group Ltd (United Kingdom [GB])
The Medicines Company (Canada)
Theranos (CA)
Theravance Inc. (CA)
Thermo Fisher Scientific (CA)
Thermo Scientific Microbiology Sdn Bhd (Malaysia)
Ventana Medical Systems Inc. (AZ)
Verinata Health, Inc. (CA)
Viracor-IBT Reference Laboratory (MO)
Wellstat Diagnostics, LLC (MD)
XDx, Inc. (CA)
Zoetis (MI)

Health Care Professions/Government

436 Medical Group - Dover Air Force Base (DE)
Aberdeen Royal Infirmary (United Kingdom [GB])
Academisch Ziekenhuis-VUB (Belgium)
ACL Laboratories (WI)
ACL Laboratories (IL)
ACM Medical Laboratory (NY)
Adams Memorial Hospital (IN)
Advanced Laboratory Services (PA)
Affiliated Laboratory, Inc. (ME)
AHS Morristown (NJ)
Akron Children's Hospital (OH)
Akron General Medical Center (OH)
Al Noor Hospital (United Arab Emirates)
Al Rahba Hospital (United Arab Emirates)
Alameda County Medical Center (CA)
Alaska Native Medical Center (AK)
Alaska Regional Hospital (AK)
Albany Medical Center Hospital (NY)
Alberta Health Services (Canada)
Alexandra Health Pte Ltd (Singapore)
Alfred I. du Pont Hospital for Children (DE)
All Children's Hospital (FL)
Allegiance Health (MI)
Alliance Community Hospital (OH)
Alpena Regional Medical Center (MI)
Alta Bates Summit Medical Center (CA)
Altru Health Systems (ND)
Alverno Clinical Laboratories, Inc. (IN)
American Association for Clinical Chemistry (DC)
American Association for Laboratory Accreditation (MD)
American Bio-Clinical Laboratories (CA)
American Medical Technologists (VA)
American Society for Clinical Pathology (IL)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

American Society for Microbiology (DC)
American Society of Phlebotomy Technicians (SC)
American Type Culture Collection (VA)
American University of Beirut Medical Ce (Lebanon)
Ampath (South Africa)
Anderson Cancer Center (TX)
Ann & Robert H. Lurie Children's Hospital of Chicago (IL)
Anne Arundel Medical Center (MD)
Anson General Hospital (Canada)
Appalachian Regional Healthcare System (NC)
Applied Proteomics Inc (CA)
Arhus Universitets Hospital (Denmark)
Arizona State Health Laboratory (AZ)
Arkansas Children's Hospital (AR)
Armed Forces Health Surveillance Center (AFHSC) (MD)
Arrowhead Regional Medical Center (CA)
Asan Medical Center (Korea, Republic of)
Asante Health System (OR)
Asia Pacific Regional - FHI360 (Thailand)
Asiri Group of Hospitals Ltd. (Sri Lanka)
ASPETAR (Qatar Orthopedic and Sports Medicine Hospital) (Qatar)
Aspirus Wausau Hospital (WI)
Associacao Das Pioneiras Sociais (Brazil)
Association of Public Health Laboratories (MD)
Atlantic Diagnostics Laboratories (PA)
Atlanticare Regional Medical Center (NJ)
Atrium Medical Center (OH)
Augusta Health (VA)
Aultman Hospital (OH)
Aultman North Canton Medical Foundation (OH)
Austin State Hospital (TX)
Avera McKennan Laboratory (SD)
AZ Sint-Lucas Hospital (Belgium)
Azienda Ospedale Di Lecco (Italy)
Banyan Biomarkers (CA)
Baptist Health Medical Center (FL)
Baptist Health System (TX)
Baptist Hospital Laboratory (FL)
Baptist Hospital of Miami (FL)
Baptist Memorial Health Care Corporation - Hospital Laboratories Works (TN)
Barnes-Jewish Hospital (VT)
Bassett Healthcare (NY)
Baton Rouge General (LA)
Baxter Regional Medical Center (AR)
Bay Area Hospital (OR)
Bay Medical Center (FL)
BayCare Health System (FL)
Bayfront Medical Center (FL)
Bayhealth Medical Center-Kent General Hospital (DE)
Baylor Health Care System (TX)
Baystate Medical Center (MA)
B.B.A.G. Ve U. AS., Duzen Laboratories (Turkey)
BC Centre for Disease Control (Canada)
Beaver Dam Reference Lab (WI)
Berlin Memorial Hospital (WI)
Berwick Hospital Center (PA)
Beth Goldstein Consultant (PA)
Beth Israel Deaconess Medical Center (MA)
Beth Israel Medical Center (NY)
Bethesda Memorial Hospital (FL)
Billings Clinic (MT)
Bio-Reference Laboratories (NJ)
Biodesign Institute at ASU (AZ)
Biothera, The Immune Health Company (MN)
Blanchard Valley Hospital (OH)
Blount Memorial Hospital (TN)
Blue Mountain Health System (PA)
Bon Secours Health Partners (VA)
Bon Secours Hospital (Ireland)
Boyce & Bynum Pathology Labs (MO)
Bozeman Deaconess Laboratory (MT)
Braintree Rehabilitation Hospital (MA)
Brant Community Healthcare System/Brant General Hospital (Canada)
Brazosport Regional Health System (TX)
Breathitt Veterinary Center, Murray State University (KY)
Bridgeport Hospital (CT)
Bristol Hospital (CT)
British Columbia Institute of Technology (Canada)
Brockville General Hospital (Canada)
Bronson Methodist Hospital (MI)
Broward General Medical Center (FL)
Brownwood Regional Medical Center (TX)
Bryan Medical Center (NE)
BSA Health System (TX)
Cadham Provincial Laboratory-MB Health (Canada)
California Pacific Medical Center (CA)
Cambridge Health Alliance (MA)
Campbellford Memorial Hospital (Canada)
Canadian Science Center for Human and Animal Health (Canada)
Canadian Society for Medical Laboratory Science (Canada)
Canberra Hospital (Australia)
Cape Fear Valley Medical Center Laboratory (NC)
Capital Health Regional Medical Center (NJ)
Capital Region Medical Center (MO)
Care Medics (Canada)
Carle Foundation Hospital (IL)
Carolinas Healthcare System (NC)
Carolinas Hospital System (SC)
Carpermor S.A. de C.V. (Mexico)
Carroll Hospital Center (MD)
Carteret General Hospital (NC)
Cary Medical Center (ME)
Castle Medical Center (HI)
Catholic Health Systems-Sisters of Charity Hospital (NY)
Catholic Medical Center (NH)
Cayuga Medical Center at Ithaca (NY)
CD Diagnostics, Inc. (PA)
Cedars-Sinai Medical Center (CA)
Cedimat Medical Center (FL)
Center for Phlebotomy Education (IN)
Centers for Disease Control and Prevention (GA)
Centers for Medicare & Medicaid Services (MD)
Central Baptist Hospital (KY)
Central Newfoundland Regional Health Center (Canada)
Central Ohio Primary Care Physicians (OH)
Central Pennsylvania Alliance Laboratory (PA)
Central Washington Hospital (WA)
Centre Hospitalier Anna-Laberge (Canada)
Centre Hospitalier Lyon SUD (France)
Ceylon Hospitals Limited (Sri Lanka)
Chaleur Regional Hospital (Canada)
Chambersburg Hospital (PA)
Champlain Valley Physicians Hospital (NY)
Chesapeake General Hospital (VA)
Chester County Hospital (PA)
Chi Solutions, Inc. (MI)
Children's Healthcare of Atlanta (GA)
Children's Hospital (AL)
Children's Hospital & Medical Center (NE)
Children's Hospital & Research Center At Oakland (CA)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

Children's Hospital Boston (MA)
Children's Hospital of Central California (CA)
Children's Hospital of Philadelphia (PA)
Children's Hospitals and Clinics (MN)
Children's Medical Center (TX)
Childrens Hospital - Kings Daughters (VA)
Childrens Hospital Los Angeles (CA)
Childrens Hospital of Wisconsin (WI)
Chino Valley Medical Center (CA)
Christiana Care Health Services (DE)
CHU-St. Justine (Canada)
Cibola General Hospital (NM)
Cincinnati Children's Hospital Medical Center (OH)
Citizens Memorial Hospital (MO)
City of Hope National Medical Center (CA)
City of Milwaukee Health Department (WI)
Cleveland Clinic (OH)
Clifton Fine Hospital (NY)
Clinica Hospital San Fernando (Panama)
Clinical Hospital Merkur (Croatia/Hrvatska)
CLMA (IL)
COLA (MD)
College of American Pathologists (IL)
College of Physicians and Surgeons of Alberta (Canada)
College of Physicians and Surgeons of Saskatchewan (Canada)
College of the North Atlantic (Canada)
College of Veterinary Medicine, Auburn University (AL)
Collingwood General & Marine Hospital (Canada)
Colorado State University (CO)
Columbia Memorial Hospital (NY)
Columbia Memorial Hospital (OR)
Commonwealth of Kentucky (KY)
Commonwealth of Virginia (DCLS) (VA)
Community College of Rhode Island-Flanagan Campus (RI)
Community Foundation of Northwest Indiana: Community Hospital (IN)
Community Hospital of the Monterey Peninsula (CA)
Community Hospitals of Williams County (OH)
Community Medical Center (MT)
Complexe Hospitalier de la Sagamie (Canada)
CompuNet Clinical Laboratories (OH)
Concord Hospital (NH)
Coney Island Hospital (NY)
Consultants Laboratory of WI LLC (WI)
Contra Costa Regional Medical Center (CA)
Cook Children's Medical Center (TX)
Cooper University Hospital (NJ)
Counties Manukau District Health Board, Middlemore Hospital (New Zealand)
Covenant Medical Center (TX)
Crozer-Chester Medical Center (PA)
CSSS Papineau/Hopital de Papineau (Canada)
CSSS St-Jerome (Canada)
Curry General Hospital (OR)
Dameron Hospital Association (CA)
Danat Al Emarat, Women and Children's Hospital (United Arab Emirates)
Darwin Health Library, NT Dept. of Health (Australia)
Daviess Community Hospital (IN)
Dayton Children's Medical Center (OH)
Deaconess Hospital Laboratory (IN)
Dean Medical Center (WI)
Delano Regional Medical Center/ Laboratory (CA)
Delaware Public Health Laboratory (DE)
Delnor Community Hospital (IL)
Denver Health Medical Center (CO)
Department of Veterans Affairs (DC)
DHHS NC State Lab of Public Health (NC)
Diagnostic Accreditation Program (Canada)
Diagnostic Center for Population & Animal Health (MI)
Diagnostic Laboratory Medicine, Inc. (MA)
Diagnostic Laboratory Services, Inc. (HI)
Diagnostic Medicine Services (Iceland)
Diagnostic Services of Manitoba (Canada)
Dialysis Clinic, Inc. Laboratory (TN)
Dimensions Healthcare System Prince George's Hospital Center (MD)
DMC University Laboratories (MI)
Docro, Inc. (CT)
Doctor's Data, Inc. (IL)
Dominican University of California (CA)
Door County Medical Center (WI)
Dr Sulaiman Al Habib Medical Group (Saudi Arabia)
Driscoll Children's Hospital (TX)
Drug Scan Inc. (PA)
DuBois Regional Medical Center (PA)
DUHS Clinical Laboratories (NC)
Duke University Medical Center (NC)
Dynacare Laboratory (WI)
DynaLIFE (Canada)
East Georgia Regional Medical Center (GA)
East Houston Regional Medical Center (TX)
East Texas Medical Center (ETMC) Henderson (TX)
East Texas Medical Center-Pittsburg (TX)
East Texas Medical Center - Tyler (TX)
Eastern Health Pathology (Australia)
Eastern Ontario Regional Laboratory Association (EORLA) (Canada)
Easton Hospital (PA)
Edgerton Hospital & Health Services (WI)
Edmonds Community College (WA)
Edward Hospital (IL)
Effingham Hospital (GA)
Emerson Hospital Laboratory (MA)
Emory University Hospital (GA)
Ephrata Community Hospital (PA)
Erie County Medical Center Corporation (NY)
Erlanger Health Systems (TN)
ESCMID (Switzerland)
Evangelical Community Hospital (PA)
Excela Health Latrobe Hospital (PA)
Exempla Lutheran Medical Center (CO)
Fairfax County Health Department (VA)
Fauquier Hospital (VA)
Fayette County Memorial Hospital (OH)
Federal Medical Center (MN)
Federal Medical Center Lexington (KY)
Firelands Regional Medical Center (OH)
Fisher-Titus Memorial Hospital (OH)
Fletcher Allen Health Care (VT)
Fleury S.A. (Brazil)
Floyd Memorial Hospital (IN)
Forrest General Hospital (MS)
Fort Defiance Indian Hospital (AZ)
Franklin Memorial Hospital (ME)
Fresno Community Hospital & Medical Center (CA)
Fundacao Faculdade de Medicina (Brazil)
Fundacion Mexicana Para la Salud Capitulo Peninsular A.C. (Mexico)
Gamma-Dynacare Laboratories (Canada)
Garden City Hospital (MI)
Genesis Healthcare System (OH)
Genesis Medical Center (IL)
Genome DX (Canada)
Genova Diagnostic Laboratory (NC)
George Mason University (VA)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

German Society of Allergy and Clinical Immunology (DGAKI) (Germany)
Ghent University Hospital (Belgium)
Glasgow Royal Infirmary (United Kingdom [GB])
Good Samaritan Hospital (IN)
Good Samaritan Hospital Medical Center (NY)
Grana S.A. (TX)
Grand River Hospital (Canada)
Grays Harbor Community Hospital (WA)
Great Plains Regional Med. Ctr. (NE)
Greater Lowell Pediatrics (MA)
Grey Bruce Regional Health Center (Canada)
Group Health Cooperative (WA)
Grove City Medical Center (PA)
Guelph General Hospital (Canada)
Gunnison Valley Hospital (CO)
Guthrie Clinic Laboratories (PA)
H. Lee Moffitt Cancer Center (FL)
Halton Healthcare Services (Canada)
Hamad Medical Corp-DLMP LAB QM (Qatar)
Hamilton Regional Laboratory Medicine Program - St. Joseph's (Canada)
Hannibal Regional Hospital (MO)
Hanover General Hospital (PA)
Hardin Memorial Hospital (KY)
Hardy Diagnostics (CA)
Harford Memorial Hospital (MD)
Harris Methodist HEB Hospital (TX)
Harris Methodist Hospital Southwest (TX)
Hartford Hospital (CT)
Hawaii State Hospital (HI)
HCA (TN)
Healdsburg District Hospital (CA)
Health City Cayman Islands (Cayman Islands)
Health Canada (Canada)
Health Network Lab (PA)
Health Sciences North (Canada)
Heartland Health (MO)
Helen Hayes Hospital (NY)
Hendrick Regional Laboratory (TX)
Hendricks Regional Health (IN)
Hendry Regional Medical Center (FL)
Henry Ford Hospital (MI)
Henry M. Jackson Foundation for the Advancement of Military Medicine-MD (MD)
Henry M. Jackson Foundation-Brook Army Medical Ctr (BAMC) (TX)
Hera General Hospital (Saudi Arabia)
Hiawatha Community Hospital (KS)
Highlands Medical Center (AL)
Hill Country Memorial Hospital (TX)
Hillcrest Medical Center (OK)
Hoag Memorial Hospital Presbyterian (CA)
Holstebro Hospital (Denmark)
Holy Name Hospital (NJ)
Holy Redeemer Hospital & Medical Center (PA)
Holy Spirit Hospital (PA)
Holzer Health System (OH)
Hong Kong Accreditation Service Innovation and Technology Commission (Hong Kong)
Hong Kong Sanatorium & Hospital (Hong Kong)
Hopital Charles Lemoyne (Canada)
Hopital Cite de La Sante De Laval (Canada)
Hopital de Granby-CSSS Haute-Yamaska (Canada)
Hopital du Haut-Richelieu (Canada)
Hopital Maisonneuve-Rosemont (Canada)
Hopital Santa Cabrini Ospedale (Canada)
Horizon Health Network (Canada)
Hospital Albert Einstein (Brazil)
Hospital Italiano Laboratorio Central (Argentina)
Hospital Sacre-Coeur de Montreal (Canada)
Hotel Dieu Grace Hospital Library (Canada)
Houston Medical Center (GA)
Hunt Regional Healthcare (TX)
Hunterdon Medical Center (NJ)
Huntington Memorial Hospital (CA)
Huntsville Memorial Hospital (TX)
Hutchinson Clinic, P.A. (KS)
Hutt Valley Health District Health Board (New Zealand)
IDEXX Reference Laboratories (Canada)
Imelda Hospital (Belgium)
Indiana University Health Bloomington Hospital (IN)
Industrial Technology Research Institute (ITRI) (Taiwan)
INEI-ANLIS Dr. C. G. Malbráin (Argentina)
Ingalls Hospital (IL)
Institut National de Sante Publique du Quebec (Canada)
Institute Health Laboratories (PR)
Institute of Tropical Medicine Dept. of Clinical Sciences (Belgium)
Institute of Veterinary Bacteriology (Switzerland)
Integrated BioBank (Luxembourg)
Integrated Regional Laboratories (HCA) (FL)
Interior Health (Canada)
Intermountain Health Care Lab Services (UT)
International Accreditation New Zealand (New Zealand)
International Federation of Clinical Chemistry (Italy)
International Health Management Associates, Inc. (IL)
Iredell Memorial Hospital (NC)
Italian Society of Clinical Biochemistry and Clinical Molecular Biology (Italy)
IU Health Bedford, Inc. (IN)
Jackson County Memorial Hospital (OK)
Jackson Health System (FL)
Jackson Hospital & Clinic, Inc. (AL)
Jackson Purchase Medical Center (KY)
Jameson Memorial Hospital (PA)
Jefferson Memorial Hospital (WV)
Jefferson Regional Medical Center (PA)
Jennings American Legion Hospital (LA)
Jessa Ziekenhuis VZW (Belgium)
John D. Archbold Hospital (GA)
John F. Kennedy Medical Center (NJ)
John H. Stroger, Jr. Hospital of Cook County (IL)
Johns Hopkins Medical Institutions (MD)
Johnson City Medical Center Hospital (TN)
Jonathan M. Wainwright Memorial Veterans Affairs Medical Center (WA)
Jones Memorial Hospital (NY)
Jordan Valley Community Health Center (MO)
JPS Health Network (TX)
Kaiser Medical Laboratory (HI)
Kaiser Permanente (GA)
Kaiser Permanente (MD)
Kaiser Permanente Colorado (CO)
Kaiser Permanente Medical Care (CA)
Kaleida Health Center for Laboratory Medicine (NY)
Kalispell Regional Medical Center (MT)
Kansas Department of Health & Environment (KS)
Kansas State University (KS)
Kaohsiun Chang Gung Memorial Hospital (Taiwan)
Karmanos Cancer Institute (MI)
Karolinska University Hospital (Sweden)
KCHL St. Elisabeth Hospital (Netherlands)
Keck Hospital of USC (CA)
Keelung Chang Gung Memorial Hospital (Taiwan)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

Kenora-Rainy River Regional Laboratory Program (Canada)
Kenya Medical Laboratory Technicians and Technologists Board (KMLTTB)
Kindred Healthcare (KY)
King Abdulaziz Hospital (Saudi Arabia)
King Fahad Specialist Hospital-Dammam, K.S.A. (Saudi Arabia)
King Faisal Specialist Hospital & Research Center (Saudi Arabia)
King Hussein Cancer Center (Jordan)
Kingston General Hospital (Canada)
KK Women's & Children's Hospital (Singapore)
Kuwait Cancer Control Center (Kuwait)
La Rabida Childrens Hospital (IL)
Lab Medico Santa Luzia LTDA (Brazil)
LABIN (Costa Rica)
Labor Stein + Kollegen (Germany)
Laboratoire National de Sante Publique (Haiti)
Laboratorio Bueso Arias (Honduras)
Laboratorio Clinico Amadita P. de Gonzales S.A. (DR)
Laboratorio de Referencia (FL)
Laboratorio Medico De Referencia (Colombia)
Laboratory Alliance of Central New York (NY)
Laboratory for Medical Microbiology and Infectious Diseases (Netherlands)
Laboratory Medicin Dalarna (Sweden)
Laboratory of Clinical Biology Ziekenhuis Oost-Limburg (ZOL) (Belgium)
LabPlus Auckland District Health Board (New Zealand)
Labrador Grenfell Health (Canada)
LAC/USC Medical Center (CA)
Lafayette General Medical Center (LA)
Lahey Hospital & Medical Center (MA)
Lake Charles Memorial Hospital (LA)
Lakeland Regional Laboratories (MI)
Lakeland Regional Medical Center (FL)
Lamb Healthcare Center (TX)
Lancaster General Hospital (PA)
Lanier Health Services (AL)
Lawrence and Memorial Hospitals (CT)
LeBonheur Children's Hospital (TN)
Legacy Laboratory Services (OR)
Leiden University Medical Center (Netherlands)
Lewis-Gale Medical Center (VA)
LewisGale Hospital Montgomery (VA)
Lexington Medical Center (SC)
Licking Memorial Hospital (OH)
LifeCare Medical Center (MN)
Lithuanian Society of Laboratory Medicine (Lithuania)
Little Company of Mary Hospital (IL)
Littleton Regional Healthcare (NH)
Lodi Health (CA)
Loma Linda University Medical Center (LLUMC) (CA)
London Health Sciences Center (Canada)
Long Beach Memorial Medical Center-LBMMC (CA)
Long Island Jewish Medical Center (NY)
Longmont United Hospital (CO)
Louisiana Office of Public Health Laboratory (LA)
Louisiana State University Medical Ctr. (LA)
Lower Mainland Laboratories (Canada)
Loyola University Medical Center (IL)
Luminex Corporation (WI)
Lutheran Hospital of Indiana Inc. (IN)
Lynchburg General (VA)
Lyndon B. Johnson General Hospital (TX)
MA Dept. of Public Health Laboratories (MA)
Mackenzie Health (Canada)
Magnolia Regional Health Center (MS)
Main Line Clinical Laboratories, Inc. Lankenau Hospital (PA)
Mammoth Hospital Laboratory (CA)
Margaret Mary Community Hospital (IN)
Margaret R. Pardee Memorial Hospital (NC)
Maria Parham Medical Center (NC)
Mariaziekenhuis vzw (Belgium)
Marion County Public Health Department (IN)
Marshall Medical Center South (AL)
Marshfield Clinic (WI)
Martha Jefferson Hospital (VA)
Martha's Vineyard Hospital (MA)
Martin Luther King, Jr./Drew Medical Center (CA)
Martin Memorial Health Systems (FL)
Mary Black Memorial Hospital (SC)
Mary Greeley Medical Center (IA)
Mary Hitchcock Memorial Hospital (NH)
Mary Washington Hospital (VA)
Massachusetts General Hospital (MA)
Mater Health Services - Pathology (Australia)
Maury Regional Hospital (TN)
Mayo Clinic (MN)
McAllen Medical Center (TX)
McCullough-Hyde Memorial Hospital (OH)
MCG Health (GA)
McGill University Health Center (Canada)
MCN Healthcare (CO)
MD Tox Laboratoires (CA)
Meadows Regional Medical Center (GA)
Med Health Services Laboratory (PA)
Medecin Microbiologiste (Canada)
Media Lab, Inc. (GA)
Medical Center Hospital (TX)
Medical Center of Central Georgia (GA)
Medical Centre Ljubljana (Slovenia)
Medical College of Virginia Hospital (VA)
Medical University Hospital Authority (SC)
Medical, Laboratory & Technology Consultants, LLC (DC)
Medlab Ghana Ltd. (Ghana)
Memorial Health System (CO)
Memorial Hermann Healthcare System (TX)
Memorial Hospital (PA)
Memorial Hospital of Texas County (OK)
Memorial Hospital of Union City (OH)
Memorial Medical Center (IL)
Memorial Regional Hospital (FL)
Memorial Sloan Kettering Cancer Center (NY)
Menonnite General Hospital (PR)
Mercy Franciscan Mt. Airy (OH)
Mercy Hospital (MN)
Mercy Hospital (IA)
Mercy Hospital of Franciscan Sisters (IA)
Mercy Hospital of Tiffin (OH)
Mercy Hospital St. Louis (MO)
Mercy Integrated Laboratories /Mercy St. Vincent (OH)
Mercy Medical Center (CA)
Mercy Medical Center (MD)
Mercy Medical Center (IA)
Mercy Medical Center (OH)
Mercy Regional Medical Center (OH)
Meritus Medical Laboratory (MD)
Methodist Dallas Medical Center (TX)
Methodist Healthcare (TN)
Methodist Hospital (TX)
Methodist Hospital Pathology (NE)
Methodist Sugarland Hospital (TX)
MetroHealth Medical Center (OH)
Metropolitan Medical Laboratory (IL)
Michigan Department of Community Health (MI)
Microbial Research, Inc. (CO)
MicroPath Laboratories (FL)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

Mid America Clinical Laboratories (IN)
Mid Coast Hospital (ME)
Middelheim General Hospital (Belgium)
Middlesex Hospital (CT)
Midland Memorial Hospital (TX)
Midwestern Regional Medical Center (IL)
Mile Bluff Medical Center/Hess Memorial Hospital (WI)
Milford Regional Hospital (MA)
Minneapolis Community and Technical College (MN)
Minneapolis Medical Research Foundation (MN)
Minnesota Department of Health (MN)
MiraVista Diagnostics (IN)
Mission Hospitals Laboratory (NC)
Mississippi Baptist Medical Center (MS)
Mississippi Public Health Laboratory (MS)
Missouri State Public Health Laboratory (MO)
MolecularMD (OR)
Monadnock Community Hospital (NH)
Monongahela Valley Hospital (PA)
Monongalia General Hospital (WV)
Montana Department of Public Health and Human Services (MT)
Montefiore Medical Center (NY)
Morehead Memorial Hospital (NC)
Mount Nittany Medical Center (PA)
Mt. Sinai Hospital (Canada)
Mt. Sinai Hospital - New York (NY)
Mt. Sinai Hospital Medical Center (IL)
MultiCare Health Systems (WA)
Munson Medical Center (MI)
Muskoka Algonquin Healthcare (Canada)
Nacogdoches Memorial Hospital (TX)
Nanticoke Memorial Hospital (DE)
Nash General Hospital/Laboratory (NC)
National Cancer Institute (MD)
National Cancer Institute, CCR, LP (MD)
National Directorate for Medical Assistance (DNAM) (Mozambique)
National Food Institute Technical University of Denmark (Denmark)
National Health Laboratory Service C/O F&M Import & Export Services (South Africa)
National Heart Institute (Institut Jantung Negra) (Malaysia)
National Institute of Health-Maputo, Mozambique (Mozambique)
National Institute of Standards and Technology (MD)
National Pathology Accreditation Advisory Council (Australia)
National Society for Histotechnology, Inc. (MD)
National University Hospital (Singapore) Pte Ltd (Singapore)
National University of Ireland, Galway (NUIG) (Ireland)
National Veterinary Institute (Sweden)
Nationwide Children's Hospital (OH)
Naval Hospital Lemoore (CA)
NB Department of Health (Canada)
Nebraska LabLine (NE)
Netlab SA (Ecuador)
New Brunswick Community College (Canada)
New Brunswick Provincial Veterinary Laboratory (Canada)
New Dar Al Shifa Hospital - Kuwait (Kuwait)
New England Baptist Hospital (MA)
New Hampshire Public Health Labs. (NH)
New Hanover Regional Medical Center (NC)
New Lexington Clinic (KY)
New London Hospital (NH)
New Medical Centre Hospital (United Arab Emirates)
New York City Department of Health and Mental Hygiene (NY)
New York Eye and Ear Infirmary (NY)
New York Presbyterian Hospital (NY)
New York State Department of Health (NY)
New Zealand Blood Service (New Zealand)
Newark Beth Israel Medical Center (NJ)
Newborn Metabolic Screening Program/ Alberta Health Services (Canada)
Newman Regional Health (KS)
Niagara Health System (Canada)
NICL Laboratories (IL)
Ninewells Hospital and Medical School (United Kingdom [GB])
NorDx - Scarborough Campus (ME)
North Bay Regional Health Center (Canada)
North Carolina Baptist Hospital (NC)
North Colorado Medical Center (CO)
North Dakota Department of Health (ND)
North District Hospital (China)
North Kansas City Hospital (MO)
North Oaks Medical Center (LA)
North Shore Hospital Laboratory (New Zealand)
North Shore Medical Center (MA)
North Shore-Long Island Jewish Health System Laboratories (NY)
Northeast Georgia Health System (GA)
Northfield Hospital & Clinics (MN)
Northside Hospital (GA)
Northside Medical Center (OH)
Northumberland Hills Hospital (Canada)
Northwest Arkansas Pathology Associates (AR)
Norton Healthcare (KY)
Nova Scotia Association of Clinical Laboratory Managers (Canada)
Nova Scotia Community College (Canada)
NSW Health Pathology (Australia)
NSW Health Pathology, Sydney South West Pathology Service (Australia)
NTD Laboratories Inc (NY)
NW Physicians Lab (WA)
Oakton Community College (IL)
Ochsner Clinic Foundation (LA)
Oconee Memorial Hospital (SC)
Octapharma Plasma (NC)
Odense University Hospital (Denmark)
Office of Medical Services Laboratory (DC)
Ohio Department of Health Lab (OH)
Ohio State University Hospitals (OH)
Oklahoma Heart Hospital, LLC (OK)
Oklahoma State University: Center for Health Sciences (OK)
Olive View-UCLA Medical Center (CA)
Olmsted Medical Center Laboratory (MN)
Oneida Healthcare Center (NY)
Ontario Medical Association Quality Management Program-Laboratory Service (Canada)
Onze Lieve Vrouwziekenhuis (Belgium)
Orange County Community College (NY)
Orange Park Medical Center (FL)
Ordre Professionnel Des Technologistes Medicaux Du Quebec (Canada)
Oregon Health and Science University (OR)
Oregon Public Health Laboratory (OR)
Orillia Soldiers Memorial Hospital (Canada)
Orlando Health (FL)
OSF - Saint Anthony Medical Center (IL)
OSU Veterinary Diagnostic Laboratory (OR)
OU Medical Center (OK)
Overlake Hospital Medical Center (WA)
Ozarks Medical Center (MO)
PA Veterinary Laboratory (PA)
Pacific Diagnostic Laboratories (CA)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

Palmetto Baptist Medical Center (SC)
Palmetto Health Baptist Easley (SC)
Palo Alto Medical Foundation (CA)
Park Nicollet Methodist Hospital (MN)
Parkview Adventist Medical Center (ME)
Parkview Health Laboratories (IN)
Parkwest Medical Center (TN)
Parrish Medical Center (FL)
Pathgroup (TN)
Pathlab (IA)
Pathology Associates Medical Lab. (WA)
PathWest Laboratory Medicine WA (Australia)
Pavia Hospital Santurce (PR)
PeaceHealth Laboratories (OR)
Peninsula Regional Medical Center (MD)
Penn State Hershey Medical Center (PA)
Pennsylvania Dept. of Health (PA)
Pennsylvania Hospital (PA)
Peoria Tazewell Pathology Group, P.C. (IL)
PEPFAR President's Emergency Plan for AIDS Relief: PEPFAR Nigeria: Medical Laboratory Sciences Council of Nigeria
PEPFAR President's Emergency Plan for AIDS Relief: PEPFAR Tanzania: Centers for Disease Control and Prevention - Tanzania
PEPFAR President's Emergency Plan for AIDS Relief: PEPFAR Tanzania: Ministry of Health and Social Welfare - Tanzania
PEPFAR President's Emergency Plan for AIDS Relief: PEPFAR Zambia: Centers for Disease Control and Prevention - Zambia
PEPFAR President's Emergency Plan for AIDS Relief: PEPFAR Zambia: Ministry of Health - Zambia
PerkinElmer Health Sciences, Inc. (SC)
Peterborough Regional Health Centre (Canada)
PHIA Project, NER (CO)
Phlebotomy Training Specialists (CA)
Phoenix Children's Hospital (AZ)
Phoenixville Hospital (PA)
PHS Indian Hospital (MN)
Physicians Choice Laboratory Services (NC)
Physicians Laboratory & SouthEast Community College (NE)
Physicians Preferred Laboratory (TX)
Placer County Public Health Laboratory (CA)
Portneuf Medical Center (ID)
Poudre Valley Hospital (CO)
Prairie Lakes Hospital (SD)
Presbyterian/St. Luke's Medical Center (CO)
Preventive Medicine Foundation (Taiwan)
Prince of Wales Hospital (Hong Kong)
Princess Margaret Hospital (Hong Kong)
Proasecal LTD (Colombia)
ProMedica Laboratory Toledo Hospital (OH)
Providence Alaska Medical Center (AK)
Providence Everett Medical Center (WA)
Providence Health Services, Regional Laboratory (OR)
Providence Hospital (AL)
Providence St. Mary Medical Center (WA)
Provista Diagnostics (AZ)
Public Health Ontario (Canada)
Pullman Regional Hospital (WA)
Queen Elizabeth Hospital (Canada)
Queen Elizabeth Hospital (China)
Queensland Health Pathology Services (Australia)
Quest - A Society for Adult Support and Rehabilitation (Canada)
Quinte Healthcare Corporation - Belleville General (Canada)
Quintiles Laboratories, Ltd. (United Kingdom [GB])
Ramathibodi Hospital (Thailand)
Range Regional Health Services (Fairview Range) (MN)
Rapides Regional Medical Center (LA)
RCPA Quality Assurance Programs Pty Limited (Australia)
Redlands Community Hospital (CA)
Regina Qu'Appelle Health Region (Canada)
Regional Laboratory of Public Health (Netherlands)
Regional Medical Laboratory, Inc. (OK)
Rehoboth McKinley Christian Health Care Services (NM)
Renown Regional Medical Center (NV)
Research Institute of Tropical Medicine (Philippines)
Rhode Island Hospital (RI)
Rice Memorial Hospital (MN)
Ridgeview Medical Center (MN)
Riverside Community Hospital (CA)
Riverside Health System (VA)
Riverside Medical Center (IL)
Robert Wood Johnson University Hospital (NJ)
Robert Wood Johnson University Hospital Rahway (NJ)
Rochester General Hospital (NY)
Roger Williams Medical Center (RI)
Roper St. Francis Healthcare (SC)
Ross University School of Veterinary Medicine (Saint Kitts and Nevis)
Roswell Park Cancer Institute (NY)
Royal Hobart Hospital (Australia)
Royal Victoria Hospital (Canada)
Rush Copley Medical Center (IL)
Rush Health Systems (MS)
Russellville Hospital (AL)
SA Pathology at Women's and Children's Hospital (Australia)
Sacred Heart Hospital (WI)
Sacred Heart Hospital (FL)
Saddleback Memorial Medical Center (CA)
Saint Francis Hospital & Medical Center (CT)
Saint Francis Medical Center (IL)
Saint Mary's Regional Medical Center (NV)
Salem Hospital (OR)
Salisbury University (MD)
Samkwang Medical Laboratory (Korea, Republic of)
Sampson Regional Medical Center (NC)
Samsung Medical Center (Korea, Republic of)
San Angelo Community Medical Center (TX)
San Francisco General Hospital-University of California San Francisco (CA)
San Jose State University (CA)
San Juan Regional Medical Group (NM)
Sanford Health (ND)
Sanford USD Medical Center (SD)
Santa Clara Valley Health & Hospital Systems (CA)
Sarasota Memorial Hospital (FL)
Saratoga Hospital (NY)
SARL Laboratoire Caron (France)
Saskatchewan Disease Control Laboratory (Canada)
Saskatoon Health Region (Canada)
Saudi Aramco Medical (TX)
SC Department of Health and Environmental Control (SC)
Schneider Regional Medical Center (Virgin Islands [USA])
Scientific Institute of Public Health (Belgium)
Scott & White Memorial Hospital (TX)
Scripps Health (CA)
Scuola Di Specializzaaione- University Milano Bicocca (Italy)
Seattle Cancer Care Alliance (WA)
Seattle Children's Hospital/Children's Hospital and Regional Medical Center (WA)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

Sentara Healthcare (VA)
Sentinel CH SpA (Italy)
Seoul National University Hospital (Korea, Republic of)
Seton Healthcare Network (TX)
Seton Medical Center (CA)
Shands Jacksonville (FL)
Shanghai Centre for Clinical Laboratory (China)
Sharon Regional Health System (PA)
Sharp Health Care Laboratory Services (CA)
Shiel Medical Laboratory Inc. (NY)
Shore Memorial Hospital (NJ)
Shriners Hospitals for Children (OH)
Silliman Medical Center (Philippines)
SIMeL (Italy)
Singapore General Hospital (Singapore)
Singulex (CA)
Slidell Memorial Hospital (LA)
SMDC Clinical Laboratory (MN)
Sociedad Espanola de Bioquimica Clinica y Patologia Molec. (Spain)
Sociedade Brasileira de Analises Clinicas (Brazil)
Sociedade Brasileira de Patologia Clinica (Brazil)
Sonora Regional Medical Center (CA)
South Bay Hospital (FL)
South Bend Medical Foundation (IN)
South Bruce Grey Health Centre (Canada)
South County Hospital (RI)
South Dakota State Health Laboratory (SD)
South Eastern Area Laboratory Services (Australia)
South Miami Hospital (FL)
South Peninsula Hospital (AK)
South West Medical Center (KS)
Southeast Alabama Medical Center (AL)
SouthEast Alaska Regional Health Consortium (SEARHC) (AK)
Southern Health Care Network (Australia)
Southern Hills Medical Center (TN)
Southwest General Health Center (OH)
Southwestern Regional Medical Center (OK)
Sparrow Hospital (MI)
Speare Memorial Hospital (NH)
Spectra East (NJ)
St Elizabeth Hospital (WI)
St Rose Dominican Hospital (AZ)
St. Agnes Healthcare (MD)
St. Anthony Hospital (OK)
St. Anthony Shawnee Hospital (OK)
St. Antonius Ziekenhuis (Netherlands)
St. Barnabas Medical Center (NJ)
St. Clair Hospital (PA)
St. David's Medical Center (TX)
St. David's South Austin Hospital (TX)
St. Elizabeth Community Hospital (CA)
St. Elizabeth's Medical Center (NY)
St. Eustache Hospital (Canada)
St. Francis Hospital (SC)
St. Francis Hospital & Health Centers (NY)
St. Francis Medical Center (LA)
St. John Hospital and Medical Center (MI)
St. John's Hospital (IL)
St. John's Hospital (WY)
St. John's Hospital & Health Center (CA)
St. John's Regional Health Center (MO)
St. Joseph Health Center (MO)
St. Joseph Health System (CA)
St. Joseph Hospital (NH)
St. Joseph Medical Center (TX)
St. Joseph Mercy - Oakland (MI)
St. Joseph Regional Health Center (TX)
St. Joseph's Hospital & Medical Center (AZ)
St. Jude Children's Research Hospital (TN)
St. Jude Medical Center (CA)
St. Luke's Episcopal Hospital (TX)
St. Luke's Hospital (IA)
St. Luke's Hospital (MN)
St. Luke's Hospital (MO)
St. Luke's Hospital (PA)
St. Luke's Hospital at The Vintage (TX)
St. Luke's Medical Center (AZ)
St. Luke's Regional Medical Center (ID)
St. Mark's Hospital (UT)
St. Mary Medical Center (CA)
St. Mary Medical Center (PA)
St. Mary's Good Samaritan (IL)
St. Mary's Health Care System (GA)
St. Mary's Health Center (MO)
St. Mary's Healthcare (NY)
St. Mary's Hospital (CO)
St. Mary's Hospital (NJ)
St. Mary's Hospital (WI)
St. Michael's Hospital/Ministry Health Care (WI)
St. Nicholas Hospital (WI)
St. Peter's Bender Laboratory (NY)
St. Peter's Hospital (MT)
St. Rita's Medical Center (OH)
St. Rose Hospital (CA)
St. Tammany Parish Hospital (LA)
St. Thomas Hospital (TN)
St. Thomas-Elgin General Hospital (Canada)
St. Vincent's Medical Center (FL)
Stanton Territorial Health Authority (Canada)
Stat Veterinary Lab (CA)
State of Alabama (AL)
State of Washington Public Health Labs (WA)
Statens Serum Institut (Denmark)
Steward Norwood Hospital (MA)
Stillwater Medical Center (OK)
Stony Brook University Hospital (NY)
Stormont-Vail Regional Medical Ctr. (KS)
Strong Memorial Hospital (NY)
Sturgis Hospital (MI)
Summa Barberton Hospital (OH)
SUNY Downstate Medical Center (NY)
Susquehanna Health System (PA)
Sutter Health (CA)
Sutter Health Sacramento Sierra Region Laboratories (CA)
SV Biosystems (CA)
Swedish American Health System (IL)
Tahoe Forest Hospital (CA)
Taiwan Society of Laboratory Medicine (Taiwan)
Tallaght Hospital (Ireland)
Tampa General Hospital (FL)
Taranaki Medlab (New Zealand)
Tartu University Clinics (Estonia)
Tataa Biocenter (Sweden)
Temple University Hospital - Parkinson Pavilion (PA)
Tenet Healthcare (PA)
Tennessee Department of Health (TN)
Tewksbury Hospital (MA)
Texas A & M University (TX)
Texas Children's Hospital (TX)
Texas Department of State Health Services (TX)
Texas Health Harris Methodist Hospital Fort Worth (TX)
Texas Health Presbyterian Hospital Dallas (TX)
Texas Scottish Rite Hospital for Children (TX)
The Charlotte Hungerford Hospital (CT)
The Cheshire Medical Center (NH)
The Children's Mercy Hospital (MO)
The Doctor's Clinic (OR)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

The Good Samaritan Hospital (PA)
The Hospital for Sick Children (Canada)
The Korean Society for Laboratory Medicine (Republic of Korea)
The Michener Institute for Applied Health Sciences (Canada)
The Naval Hospital of Jacksonville (FL)
The Nebraska Medical Center (NE)
The Norwegian Institute of Biomedical Science (Norway)
The Permanente Medical Group, Inc. (CA)
The University of Texas Medical Branch (TX)
The University of Tokyo (Japan)
Thomas Jefferson University Hospital, Inc. (PA)
Thomas Memorial Hospital (WV)
Timmins and District Hospital (Canada)
Torrance Memorial Medical Center (CA)
Touro Infirmary (LA)
Tri-Cities Laboratory (WA)
TriCore Reference Laboratories (NM)
Trillium Health Partners Credit Valley Hospital (Canada)
Trinity Medical Center (AL)
Trinity Muscatine (IA)
Tucson Medical Center (AZ)
Tuen Mun Hospital, Hospital Authority (Hong Kong)
Tufts Medical Center (MA)
Tulane Medical Center Hospital & Clinic (LA)
Tulane University Health Sciences Center (LA)
Twin Lakes Regional Medical Center (KY)
U.S. Medical Center for Federal Prisoners (MO)
UC Davis Medical Center Department of Pathology & Laboratory Medicine (CA)
UC San Diego Health System Clinical Laboratories (CA)
UCI Medical Center (University of California, Irvine) (CA)
UCLA Medical Center (CA)
UCONN Health Center (CT)
UCSF Medical Center China Basin (CA)
UMass Memorial Medical Center (MA)
UMC of El Paso- Laboratory (TX)
UMC of Southern Nevada (NV)
Umea University Hospital (Sweden)
UNC Hospitals (NC)
United Christian Hospital (Hong Kong)
United Clinical Laboratories (IA)
United Health Services Hospital/Wilson Hospital Laboratory (NY)
United Memorial Medical Center (NY)
United States Coast Guard (NJ)
Universidad de Guadalajara (Mexico)
Universitair Ziekenhuis Antwerpen (Belgium)
University College Hospital (Ireland)
University General Hospital (TX)
University Health Network (Canada)
University Hospital (TX)
University Hospital Center Sherbrooke (CHUS) (Canada)
University Hospital of Northern BC (Canada)
University Hospitals of Cleveland (OH)
University Medical Center (TX)
University of Alabama at Birmingham (AL)
University of Alabama Hospital Laboratory (AL)
University of Arizona Medical Center (AZ)
University of Bonn (Germany)
University of California Veterinary Medical Teaching Hospital (CA)
University of Chicago Hospitals (IL)
University of Cologne Medical Center (Germany)
University of Colorado Denver, Anschutz Medical Campus (CO)
University of Colorado Hospital (CO)
University of Guelph (Canada)
University of Idaho (ID)
University of Illinois Medical Center (IL)
University of Iowa Hospitals and Clinics (IA)
University of Iowa, Hygienic Lab (IA)
University of Louisville Hospital (KY)
University of Maryland Medical System (MD)
University of Miami (FL)
University of Michigan, Department of Pathology (MI)
University of Minnesota Medical Center-Fairview (MN)
University of Missouri Hospital (MO)
University of North Carolina - Health Services (NC)
University of Oregon (OR)
University of Pennsylvania (PA)
University of Pennsylvania Health System (PA)
University of Pittsburgh Medical Center (PA)
University of Prince Edward Island Atlantic Veterinary College (Canada)
University of Rochester Medical Center (NY)
University of South Alabama Medical Center (AL)
University of Tasmania (Australia)
University of Texas Health Center (Tyler) (TX)
University of Texas Health Science Center (TX)
University of Texas Southwestern Medical Center (TX)
University of Utah Hospital & Clinics (UT)
University of Virginia Medical Center (VA)
University of Washington Medical Center (WA)
University of Wisconsin Health (WI)
UPMC Bedford Memorial (PA)
Uvalde Memorial Hospital (TX)
UZ-KUL Medical Center (Belgium)
VA (Bay Pines) Medical Center (FL)
VA (Indianapolis) Medical Center (IN)
VA (Miami) Medical Center (FL)
VA (Tampa) Hospital (FL)
VA (Tuscaloosa) Medical Center (AL)
Vail Valley Medical Center (CO)
Valley Medical Center (WA)
Vancouver Island Health Authority (SI) (Canada)
Vanderbilt University Medical Center (TN)
Vejle Hospital (Denmark)
Vernon Memorial Hospital (WI)
Via Christi Hospitals - Wichita (KS)
Vibrant America LLC (CA)
Vidant Medical Center (NC)
Virginia Mason Medical Center (WA)
Virginia Physicians, Inc. (VA)
Virtua - West Jersey Hospital (NJ)
WakeMed (NC)
Waterbury Hospital (CT)
Watson Clinic (FL)
Wayne Healthcare (OH)
Wayne Memorial Hospital (GA)
Weeneebayko General Hospital (Canada)
Weirton Medical Center (WV)
Wellstar Health Systems (GA)
Wenatchee Valley Medical Center (WA)
Wesley Medical Center (KS)
West Georgia Health Systems (GA)
West Kendall Baptist Hospital (FL)
West Shore Medical Center (MI)
West Valley Medical Center Laboratory (ID)
West Virginia University Hospitals (WV)
Westchester Medical Center (NY)
Western Healthcare Corporation (Canada)
Western Maryland Regional Medical Center (MD)
Western Missouri Medical Center (MO)
Western Reserve Hospital (OH)
Western State Hospital (VA)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

Whangarei Hospital (New Zealand)
Wheaton Franciscan Laboratories at St. Francis (WI)
Wheeling Hospital (WV)
Whitehorse General Hospital (Canada)
Whitman Hospital & Medical Center (WA)
Wickenburg Community Hospital (AZ)
William Beaumont Army Medical Center (TX)
William Osler Health Centre (Canada)
Williamson Medical Center (TN)
Winchester Hospital (MA)
Winter Haven Hospital, Inc. (FL)
Wisconsin State Laboratory of Hygiene (WI)
Women & Infants Hospital (RI)
Women's and Children's Hospital (LA)
Woodside Health Center (Canada)
World Health Organization (Switzerland)
WuXi AppTec Co., Ltd. (China)
Wyckoff Heights Medical Center (NY)
Yale New Haven Hospital (CT)
York General Health Care Services (NE)
York Hospital (PA)
Yukon-Kuskokwim Delta Regional Hospital (AK)
Yuma Regional Medical Center (AZ)

Individuals

Mohamed O Abdelhalim (Oman)
Franklin Adkinson (MD)
Park Ae Ja (Korea, Republic of)
Shana Ahmad (NY)
Fayha Ahmed (United Kingdom [GB])
Lawal Akeem (Nigeria)
Evelyn W. Akers (NY)
Ahmed M. Albarrag (Saudi Arabia)
Erika B Ammirati (CA)
Mohammed Attaelmannan (MA)
Chris Aug (CT)
Ahmed M Azaybi (Saudi Arabia)
Cary Baird (OH)
Amy Baracz (CA)
Susan Barber (NC)
Colette Batog (PA)
Joanne Becker (NY)
Nancy Behling (AZ)
Lynn Bell (IN)
Steven Bellistri (PA)
Melissa Bennett (Canada)
Dr. Lynette Y. Berkeley PhD (MD)
Ms. Lucia M. Berte MT(ASCP) SBB, DLM; CQA(ASQ) CMQ/OE (CO)
Bhaskar Bhattacharya (India)
Elma Kamari Bidkorpeh (CA)
Abbejane Blair (MA)
Dennis Bleile (CA)
Ms. Susan Blonshine RRT, RPFT, FAARC (MI)
Fran Boemer (Belgium)
Elizabeth Brown (PA)
Steven Brown (OR)
Carey-Ann Burnham (MO)
Karen Bush (IN)
Donald R Callihan (MD)
Rebecca Cameron (MS)
Ms. Natalie Campbell RT (Canada)
Sheldon Campbell (CT)
Alan T. Cariski (CA)
A. Bjoern Carle (ME)
Dr. Maria Paz Carlos DVM, PhD, MBA (MD)
Eileen Carreiro-Lewandowski (MA)
Dr. Alexis Carter MD, FCAP, FASCP (GA)
Dr. Jose B. Casals (Denmark)
Ning Cegielski (WA)
Emily Chang (VA)
Mintrude Charles-Young (Canada)
Denise Chorley (CA)
W. Gregory Cooper, LLC (TX)
William A Coughlin (VT)
Pauline Cyr (Canada)
Redintor Dagos (Philippines)
Dr. Jeff Dahlen PhD (CA)
Imelda Daniel (CA)
Saffiatou Darboe (Gambia)
Ms. Arlene Darmanie MS (Trinidad and Tobago)
Dr. Trivikram Dasu PhD (WI)
Ms. Diana R. DeHoyos MS, MT(ASCP) (TX)
Dr. Maria del Pilar Aguinaga PhD, CLDir(NCA) (TN)
N. de Jonge (Netherlands)
Anne Delaney (AZ)
Dr. Francois Depasse PharmD, MSc (France)
Narendra Desai (CA)
Dr. Edward P. Desmond PhD (CA)
Patricia Devine (MA)
Tom Dew (PA)
Ms. Diana L. Dickson MS, RAC (PA)
Donna Downs (NV)
Dr. Sherry A. Dunbar PhD (TX)
Mr. A. Paul Durham MA (CA)
Kathleen Dwyer (TX)
Pinar Eker (Turkey)
Sahar Gamil EL-Wakil (Egypt)
Dr E Elnifro (Malta)
Paulo Enrico P. Belen (Philippines)
Mike Ero (CA)
Mr. German Esparza BSc (Colombia)
Amy F, MS (NY)
Dr. William Fales (MO)
Pilar Fernandez-Calle (Spain)
Leah Ferrier (MT)
Ms. Sue Forrest (Australia)
Marcia Foxworthy (AL)
Dr. Timothy S. Frana DVM, MS, MPH, PhD (IA)
Dr. Jeff Fuller PhD, FCCM, ABMM (Canada)
Mary Lou Gantzer (DE)
Dr. Valerio M. Genta MD (VA)
Karima Ghazzaly (TX)
Marc Goldford (IN)
Merran Govendir (Australia)
Tanya Graham (SD)
Neil Greenberg (NY)
Ann M. Gronowski (MO)
Jason Gruver (IA)
Dr. Tibor Gyorfi (GA)
John F Halsey (SC)
Dr. W. Harry Hannon PhD (GA)
Syed N Hassan (NY)
Alandria Hatcher (TX)
B. Y. Hsieh (Taiwan)
Po-Ren Hsueh (Taiwan)
Mr. Darren C. Hudach (OH)
Doreene Hyatt (CO)
Anne Igbokwe (CA)
Lugard Igharo (TX)
Cathy Trumel (France)
T. S Isbell (MO)
Dr. Megan E. Jacob PhD (NC)
Ellis Jacobs (NJ)
Carlos A. Javier (FL)
Amy Jerabek (WI)
Daniel M. Johnson (IA)
Judith Johnston (CA)
Stephen Kahn (IL)
Jiesheng Kang (MA)
Nachum Kaplan (Canada)
Mr. Bob Kaplanis PBT, MT(ASCP) (AZ)
David Kasper (Austria)
Dr. Steven C. Kazmierczak PhD, DABCC, FACB (OR)

© Clinical and Laboratory Standards Institute. All rights reserved.

Active Membership (As of 1 August 2014)

Natalie J. Kennel (CA)
Michael Kent (OH)
Mr. Klaus M. Kjoller MSc (Denmark)
William F. Koch (MD)
Denise Kramer (OH)
Teresa Kraus (OK)
Mr. Narayan Krishnaswami MS, MBA (MO)
Martin Kroll (NJ)
Jan Krouwer (MA)
Kristi Kuper (TX)
Dr. Patrick B. Kyle PhD (MS)
Giancarlo la Marca (Italy)
Michael LaFleur (MA)
Grace Largado (CA)
Professor Szu-Hee Lee MD, PhD (Australia)
Dr. Thomas J. Lenk PhD (CA)
Sarah B Leppanen (CA)
Andrew Leung (CA)
Jacob B Levine (NJ)
Ernst Lindhout (Netherlands)
Kristian Linnet (Denmark)
Yuqing Liu (China)
Philip Lively (PA)
Mark Loch (MN)
Moushumi Lodh (India)
Brian Lubbers (KS)
Darrell Lundrigan (Canada)
Dr. Raquel Yahyaoui Macias (Spain)
Dr. Roberta Madej (CA)
Edward Mahamba (NV)
Mr. David Manalan F(ASQ), CSQE, CBA (MA)
Wilma Mangan (CA)
Linda M Mann (CA)
Adrienne Manning (CT)
Kristin M Marckel (MN)
Richard A. Marlar (OK)
Barbara Masten (NM)
Christine McRoberts (Canada)
Dr. Piet Meijer PhD (Netherlands)
Jacques F Meis (Netherlands)
Vanessa D. Mercado (Puerto Rico)
Barbara Mervin (MD)
James J. Miller (KY)
Kathryne Miskavige (ND)
Ms. Barbara Mitchell (KS)
Kristine Moraski (PA)
Yaser Morgan (NY)
Ian Morrissey (Switzerland)
Mohamed Hanafy Morsy (Saudi Arabia)
Anna Murphy (NJ)
Nombuso Ndlovu (South Africa)
Vishwanand Negi (India)
Melanie O'Keefe (Australia)
Jeffrey O'Kelley (GA)
Joseph Oduor Ochieng (Kenya)
Olajumoke Oladipo (NY)
Mr. Gregory Olsen (NE)
Ms. Margaret Ordonez Smith de Danies (Colombia)
Mr. Jan Ostrup (Finland)
Dr. Elizabeth Palavecino MD (NC)
Dr. Mark G. Papich DVM, MS (NC)
Niketankumar Patel (MI)
Ketan Patel (MD)
Dr. Deborah Payne PhD (CO)
A. K. Peer (South Africa)
Armando Perez-Cardona (FL)
David S Perlin (NJ)
Linda Perryman (GA)
Philip A Poston, PhD (VA)
Dr. Mair Powell MD, FRCP, FRCPath (United Kingdom [GB])
Nicole Procise (IN)
Dr. Mathew Putzi (TX)
Dr. Qinfang Qian MD, PhD (MA)
Albert Rabinovitch (CA)
Tawni Reller (MN)
Ms. Allison Remensperger (CA)
Dr. Robert P. Rennie PhD (Canada)
Mary Rice (CO)
Hanna Ritzen (Sweden)
Jennifer Rogers (MI)
Dr. Markus Rose DVM, PhD (Germany)
Andreas Rothstein (Colombia)
Daniel Ryan (CA)
Habib Sadat Chaudaury (Bangladesh)
Dr. Linoj Samuel PhD, D(ABMM) (MI)
Dr. Leticia J. San Diego PhD (MI)
Caroline Satyadi (CA)
Theresa Schnellman (SC)
Paul Schreckenberger (IL)
Melvin Schuchardt (GA)
Kathleen Selover (NY)
Dan Shireman (KS)
Dinah Shore Myers (NC)
Dr. Venkatakrishna Shyamala PhD (MD)
Dr. Vijay K. Singu DVM, PhD (NE)
Martha Skorczewski (IL)
Jane L. Smith (TN)
Janis F. Smith, Laboratory (MD)
Anna V. Sombong (Philippines)
Oyetunji O. Soriyan (Nigeria)
Paul D Stamper (MD)
John Stelling (MA)
Jennifer Strom (TX)
Ara Tahmassian (MA)
Len Tanaka (HI)
José Eduardo Mucito Varela (Mexico)
Michelle Vanderpool (CA)
Suresh H Vazirani (India)
Ryan A. Vicente (Qatar)
Lenin Villalta (Ecuador)
Kim Walker (CA)
Megan Waller (MD)
Dr. Hui Wang PhD (China)
Peter Warn (United Kingdom [GB])
Mr. Niels Wartenberg (MN)
Markita Weaver (MA)
Mr. Marlon A. Webb (MD)
Alice S Weissfeld (TX)
Gary Wells (TX)
Eric Whitters (PA)
Dr L.A. Nilika Wijeratne (Australia)
Bernadette Wildemore (GA)
Dr. Emily S. Winn-Deen PhD (CA)
Mr. Dennis Winsten (AZ)
William W Wood (MA)
Ms. Sheila M. Woodcock ART, MBA, FCSMLS(D) (Canada)
Ginger Wooster (WI)
Dr. Ching Ching Wu DVM, PhD (IN)
Max Wu (CA)
Dr. Shangwei Wu PhD (China)
Stanley Wu (GA)
Idris Mohammed Yahaya (Nigeria)
Patty Yates (MS)
Steven York (OH)
Atsushi Yoshida (Japan)
Michelle L. Zaharik (Canada)
Dr. Marcia L. Zucker PhD (NJ)
Kimberly Zunker (TX)

© Clinical and Laboratory Standards Institute. All rights reserved.

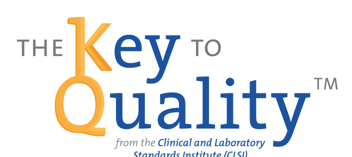
NOTES

© Clinical and Laboratory Standards Institute. All rights reserved.

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

Explore the Latest Offerings from CLSI!

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



Power Forward with this Official Interactive Guide

Fundamentals for implementing a quality management system in the clinical laboratory.

Visit the CLSI U Education Center

Where we provide the convenient and cost-effective education resources that laboratories need to put CLSI standards into practice, including webinars, workshops, and more.

Shop Our Online Products

Including eCLIPSE Ultimate Access™, CLSI's cloud-based, online portal that makes it easy to access our standards and guidelines—*anytime, anywhere.*

Find Membership Opportunities

See the options that make it even easier for your organization to take full advantage of CLSI benefits and our unique membership value.

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

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

P: 610.688.0100 **Toll Free (US):** 877.447.1888 **F:** 610.688.0700

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

PRINT ISBN 1-56238-967-X

ELECTRONIC ISBN 1-56238-968-8