EP5-A2
Vol. 24 No. 25
Replaces EP5-A
Vol. 19 No. 2

Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition

This document provides guidance for designing an experiment to evaluate the precision performance of quantitative measurement methods; recommendations on comparing the resulting precision estimates with manufacturers' precision performance claims and determining when such comparisons are valid; as well as manufacturers' guidelines for establishing claims.

A guideline for global application developed through the NCCLS consensus process.

NCCLS...

Global Consensus Standardization for Health Technologies

NCCLS is an international, interdisciplinary, nonprofit, standards-developing, and educational organization that promotes the development and use of voluntary consensus standards and guidelines within the healthcare community. It is recognized worldwide for the application of its unique consensus process in the development of standards and guidelines for patient testing and related healthcare issues. NCCLS is based on the principle that consensus is an effective and cost-effective way to improve patient testing and healthcare services.

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

PUBLICATIONS

An NCCLS document is published as a standard, guideline, or committee report.

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

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

Report A document that has not been subjected to consensus review and is released by the Board of Directors.

CONSENSUS PROCESS

The NCCLS voluntary consensus process is a protocol establishing formal criteria for:

- the authorization of a project
- the development and open review of documents
- the revision of documents in response to comments by users
- the acceptance of a document as a consensus standard or guideline.

Most NCCLS documents are subject to two levels of consensus—"proposed" and "approved." Depending on the need for field evaluation or data collection, documents may also be made available for review at an intermediate consensus level.

Proposed An NCCLS consensus document undergoes the first stage of review by the healthcare community as a proposed standard or guideline. The document should receive a wide and thorough technical review, including an overall review of its scope, approach, and utility, and a line-by-line review of its technical and editorial content.

Approved An approved standard or guideline has achieved consensus within the healthcare community. It should be reviewed to assess the utility of the final document, to ensure attainment of consensus (i.e., that comments on earlier versions have been satisfactorily addressed), and to identify the need for additional consensus documents.

NCCLS standards and guidelines represent a consensus opinion on good practices and reflect the substantial agreement by materially affected, competent, and interested parties obtained by following NCCLS's established consensus procedures. Provisions in NCCLS standards and guidelines may be more or less stringent than applicable regulations. Consequently, conformance to this voluntary consensus document does not relieve the user of responsibility for compliance with applicable regulations.

COMMENTS

The comments of users are essential to the consensus process. Anyone may submit a comment, and all comments are addressed, according to the consensus process, by the NCCLS committee that wrote the document. All comments, including those that result in a change to the document when published at the next consensus level and those that do not result in a change, are responded to by the committee in an appendix to the document. Readers are strongly encouraged to comment in any form and at any time on any NCCLS document. Address comments to the NCCLS Executive Offices, 940 West Valley Road, Suite 1400, Wayne, PA 19087, USA.

VOLUNTEER PARTICIPATION

Healthcare professionals in all specialties are urged to volunteer for participation in NCCLS projects. Please contact the NCCLS Executive Offices for additional information on committee participation.

EP5-A2
ISBN 1-56238-542-9
Volume 24 Number 25 ISSN 0273-3099

Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition

Daniel W. Tholen, M.S.
Anders Kallner, M.D., Ph.D.
John W. Kennedy (deceased)
Jan S. Krouwer, Ph.D.
Kristen Meier, Ph.D.

Abstract

NCCLS document EP5-A2, *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition* provides guidance and procedures for evaluating the precision of *in vitro* diagnostic devices and includes recommendations for manufacturers in evaluating their devices and methods when establishing performance claims. Included are guidelines for the duration, procedures, materials, data summaries, and interpretation techniques that are adaptable for the widest possible range of analytes and device complexity. The procedures are designed for manufacturers or developers of clinical laboratory measurement methods, and for users of those methods who wish to determine their own performance capabilities or to verify claims from a manufacturer. A balance is created in the document between complexity of design and formulae, and simplicity of operation.

NCCLS. *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition*. NCCLS document EP5-A2 (ISBN 1-56238-542-9). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2004.

NCCLS

Suggested Citation

(NCCLS. *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition.* NCCLS document EP5-A2 [ISBN 1-56238-542-9]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2004.)

Proposed Guideline
December 1981

Tentative Guideline
December 1982

Approved Guideline
February 1999

Approved Guideline—Second Edition
August 2004

ISBN 1-56238-542-9
ISSN 0273-3099

Committee Membership

Area Committee on Evaluation Protocols

Jan S. Krouwer, Ph.D.
Chairholder
Krouwer Consulting
Sherborn, Massachusetts

Luann Ochs, M.S.
Vice-Chairholder
Roche Diagnostics Corporation
Indianapolis, Indiana

Anders Kallner, M.D., Ph.D.
Karolinska Hospital
Stockholm, Sweden

Martin Harris Kroll, M.D.
Dallas VA Medical Center
Dallas, Texas

Jacob (Jack) B. Levine, M.B.A.
Bayer Corporation
Tarrytown, New York

Kristian Linnet, M.D., Ph.D.
Psychiatric University Hospital
Risskov, Denmark

Kristen L. Meier, Ph.D.
FDA Ctr. for Devices/Rad. Health
Rockville, Maryland

Max Robinowitz, M.D.
FDA Ctr. for Devices/Rad. Health
Rockville, Maryland

Daniel W. Tholen, M.S.
Dan Tholen Statistical Consulting
Traverse City, Michigan

Advisors

David A. Armbruster, Ph.D., DABCC, FACB
Abbott Laboratories
Abbott Park, Illinois

R. Neill Carey, Ph.D.
Peninsula Regional Medical Center
Salisbury, Maryland

Carl C. Garber, Ph.D., FACB
Quest Diagnostics, Incorporated
Teterboro, New Jersey

Patricia E. Garrett, Ph.D.
Boston Biomedica, Inc.
West Bridgewater, Massachusetts

John W. Kennedy (deceased)
Medstat Consultants
San Francisco, California

Donald M. Powers, Ph.D.
Powers Consulting Services
Pittsford, New York

Gian Alfredo Scassellati, Ph.D.
Ente Nazional Italiano Di Unificatione
Turin, Italy

Jack Zakowski, Ph.D.
Beckman Coulter, Inc.
Brea, California

Staff

Lois M. Schmidt, D.A.
Staff Liaison
NCCLS
Wayne, Pennsylvania

Donna M. Wilhelm
Editor
NCCLS
Wayne, Pennsylvania

Melissa A. Lewis
Assistant Editor
NCCLS
Wayne, Pennsylvania

Acknowledgements

NCCLS and the Area Committee on Evaluation Protocols gratefully acknowledge the invaluable contributions of the late John W. Kennedy to the development of the EP5 document. John led the original NCCLS project to develop a procedure for estimating precision, and he patiently guided EP5 through the consensus process, including the current revision of the approved document. John gave generously of his time and his talents to many other NCCLS evaluation protocols, provided statistical support for other NCCLS committees, and trained many laboratory professionals in the proper use of the statistical protocols. He will be missed.

The Area Committee on Evaluation Protocols would also like to recognize the valuable contributions of the members and advisors of the Subcommittee on Evaluation of Precision that developed the first approved edition of this guideline.

John W. Kennedy, Chairholder (deceased)
R. Neill Carey, Ph.D.
Richard B. Coolen, Ph.D.
Carl C. Garber, Ph.D.
Henry T. Lee, Jr.
Jacob B. Levine, M.B.A.
Iris M. Osberg
Stanley Bauer, M.D. (deceased)
James O. Westgard, Ph.D.
Donald M. Powers, Ph.D.

Contents

Contents (Continued)

Foreword

Current clinical chemistry literature contains numerous examples of product evaluations. Many of these use the basic concepts that are included in this guideline. While more complex and customized experimental designs have been used for both published studies and regulatory purposes in special cases, there still appears to be a strong need in the clinical community for the basic approaches to quantitative precision assessment to be described, as well as their rationales.

In order to address this need, the committee has drawn on the experience of users, representatives of industry, statisticians, chemists, laboratory personnel, regulatory authorities, and medical personnel for developing this guideline. The extremely wide variety of *in vitro* diagnostic devices currently available made it apparent that a single experimental design would not be appropriate for all devices. Therefore, this guideline has been constructed to provide primarily conceptual guidance on the duration, procedures, materials, data summaries, and interpretation techniques that would be adaptable for the widest possible range of analytes and device complexity. Illustrations of each step of the evaluation, with an example of a typical experimental design, have also been provided.

In development of this protocol, many recommendations for duration, inclusion of quality control, and methods of determining the components of precision were carefully considered. The resultant protocol creates a balance between complexity of design and formulae, and simplicity of operation. For ease of use, an appendix (Appendix C) has been included that provides guidelines for modifying the design and calculations when appropriate.

Since its publication as a tentative guideline in 1992 and then as an approved guideline in 1999, document EP5 has been widely used by device manufacturers to establish precision claims for their methods, and by laboratories to determine the performance of methods in their use. In preparing EP5-A2, the Area Committee on Evaluation Protocols sought to retain the experimental procedures while harmonizing the terminology with current international recommendations. Along with terminology changes, text has been added to distinguish different users of the document, and sections were added on Scope and Definitions, per current NCCLS document format. Comments received on the present edition are included in the comment/response summary starting on page 35. The area committee plans to broaden the scope of the next revision of this document to address more complex procedural issues and expand the statistical analysis involved in evaluating precision.

The various components of precision in EP5-A2, especially “within-laboratory precision,” could be important components of *measurement uncertainty (MU),* per current ISO usage: “a parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand.” However, EP5-A2 procedures cannot be used alone to estimate MU. There are other components for many analytes, and different ways to combine them. Fully ISO-compliant calculations of measurement uncertainty involve concepts and procedures that are beyond the scope of the current document, but may be addressed in the next edition.

A Note on Terminology

NCCLS, as a global leader in standardization, is firmly committed to achieving global harmonization in terminology wherever possible. Harmonization is a process of recognizing, understanding, and explaining differences in terms while taking steps to achieve worldwide uniformity. NCCLS 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 NCCLS, ISO, and CEN documents; and that legally required use of terms, regional usage, and different consensus timelines are all obstacles to harmonization. In light of this, NCCLS recognizes that harmonization of terms facilitates the global application of standards and deserves immediate attention. Implementation of this policy must be an evolutionary and educational process that begins with new projects and revisions of existing documents.

In order to align the usage of terminology in this document with that of ISO, the term *sample* has replaced the term s*pecimen* and the term *measuring range* has replaced the term *reportable range.* The users of EP5-A2 should understand that the fundamental meanings of the terms are similar, and to facilitate understanding, where appropriate, the terms are defined along with their ISO counterpart in the guideline's Definitions section. (See Section 4.)

The term *precision* is always used as a measure of "closeness of agreement between independent test/measurement results obtained under stipulated conditions."[1] The terms in this document are consistent with uses defined in the ISO 3534 and ISO 5725 series of standards. In these models, *repeatability* and *reproducibility* are considered to be the extreme measures of precision, with repeatability being the smallest measure (same operator, method, equipment, time, and laboratory) and reproducibility being the largest (different operator, equipment, and laboratory). All other measures of precision are "intermediate measures" and must be explicitly described. Reproducibility is not estimated in EP5-A2, since the protocol does not require multiple laboratories. All other measures of precision from EP5-A have been retained, although the term *total precision* was eliminated, because it was not clearly defined. In this document, *total precision* has been replaced by *within-laboratory* or *within-device,* depending on whether the laboratory or manufacturer is deriving the estimate.

Key Words

Evaluation protocol, experimental design, medical devices, outlier, precision, quality control

Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition

1 Scope

This document provides guidelines for experiments to evaluate the precision performance characteristics of quantitative measurement methods and devices. It includes protocols for developers of testing methods or devices, and protocols for users of these methods who wish to determine their own performance capabilities. These procedures may not be appropriate for some quantitative methods for which adequate test materials do not exist.

2 Introduction

This document is for manufacturers of *in vitro* diagnostic (IVD) devices and developers of clinical laboratory measurement methods who wish to establish the precision capabilities of their methods. It is also for the users of those methods who wish to verify the validity of performance claims, or who simply want to measure their own precision. Users of automated measurement procedures who wish only to apply a minimal protocol to verify the validity of a manufacturer's claims for precision should follow the guidance of the most current edition of NCCLS document EP15—*User Demonstration of Performance for Precision and Accuracy*. The guidelines are fully general for these situations, because they include considerations of goals for the reliability of the precision estimates.

This document also applies to laboratories that make significant modifications to current methods. When using a modification of an *in vitro* diagnostic (IVD) device or method, a user needs to verify that essential performance characteristics of the device have not changed. Comparison to original claimed precision performance may not be valid. Examples of typical modifications are the use of reagents, sample sources, calibrating or control materials, or operating procedures that are different from those stated in the manufacturer's labeling (instructions for use).

3 Standard Precautions

Because it is often impossible to know what 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 infectious agents and thus are more comprehensive than universal precautions which are intended to apply only to transmission of blood-borne pathogens. Standard and universal precaution guidelines are available from the U.S. Centers for Disease Control and Prevention (*Guideline for Isolation Precautions in Hospitals*. Infection Control and Hospital Epidemiology. CDC. 1996;17(1):53-80 and *MMWR* 1988;37:377-388). For specific precautions for preventing the laboratory transmission of all infectious agents from laboratory instruments and materials and for recommendations for the management of exposure to all infectious disease, refer to the most current edition of NCCLS document M29—*Protection of Laboratory Workers from Occupationally Acquired Infections*.

4 Definitions

Analyte – Component represented in the name of a measurable quantity; **NOTES:** a) This includes any element, ion, compound, substance, factor, infectious agent, cell, organelle, activity (enzymatic, hormonal, or immunological), or property, the presence or absence, concentrations, activity, intensity, or other characteristics of which are to be determined; b) 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)[2]; c) In the type of quantity "catalytic concentration of lactate dehydrogenase isoenzyme 1 in plasma," "lactate dehydrogenase isoenzyme 1" is the analyte (ISO 18153).[3]

Imprecision – Dispersion of independent results of measurements obtained under specified conditions.

Intermediate precision (measure) – Precision under intermediate precision conditions.

Intermediate precision conditions – Where test results or measurement results are obtained with the same method, on identical test/measurement items in the same test facility, under some different operating condition; **NOTES:** a) There are four elements to the operating conditions: time, calibration, operator, and equipment; b) 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 – Particular quantity subject to measurement (VIM93)[4]; **NOTE:** This term 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.

Measuring range//(Reportable range) – A set of values of measurands for which the error of a measuring instrument is intended to lie within specified limits; **NOTES:** a) For this document, the range of values (in units appropriate for the analyte [measurand]) over which the acceptability criteria for the method have been met; that is, where errors due to nonlinearity, imprecision, or other sources are within defined limits; b) Formerly **Reportable range** – The range of test values over which the relationship between the instrument, kit, or system's measurement response is shown to be valid.

Precision – Closeness of agreement between independent test/measurement results obtained under stipulated conditions.

Repeatability conditions – Conditions where independent test results are obtained with the same method on identical test material in the same laboratory by the same operator using the same equipment within a short interval of time (ISO 3534-1); **NOTE:** Formerly termed *within-run precision*.

Repeatability (of results of measurements) – Closeness of the agreement between results of successive measurements of the same measurand carried out under the same conditions of measurement (VIM93).

Reproducibility conditions – Conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment (ISO 5725-1).[5]

Reproducibility (of results of measurements) – Closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement (VIM93).

Run – An interval within which the trueness and precision of a testing system is expected to be stable, but cannot be greater than 24 hours or less than the frequency recommended by the manufacturer (US CFR493 February 28, 1992);[6] **NOTE:** (ISO3534-1/93-2.48) defines "run" as follows: In a series of observations of a qualitative characteristic, the occurrence of an uninterrupted series of the same attribute is called a "run."

Sample – One or more parts taken from a system, and intended to provide information on the system, often to serve as a basis for decision on the system or its production; **NOTE:** For example, a volume of serum taken from larger volume of serum.

Within-device/Within-run precision – See **Intermediate precision conditions**.

5 Symbols Used in Text

A	standard deviation of the run means
B	standard deviation of the daily means
χ^2	Chi-square statistic for testing manufacturers' claims
I	total number of days (generally 20)
J	number of runs within-day (generally 2)
MD	mean square for days
ME	mean square for within-run (error)
MR	mean square for runs
R	total number of runs (1+degrees of freedom for S^2_r)
S_{dd}	estimate of between-day standard deviation
S_r	estimate of repeatability standard deviation (within-run precision)
S_{rr}	estimate of between-run standard deviation
S_T	estimate of within-device or within-laboratory precision standard deviation
σ_r	performance claim repeatability standard deviation (within-run precision)
σ_T	manufacturer's claim of total standard deviation or medically required standard deviation
T	1+degrees of freedom for S_T
X_{ijk}	result for run j on day i (result of replicate k on run j on day i; generally k = 1 or 2)
$\overline{X}_{i_1\bullet}$	average result of the replicates for run 1, day I
$\overline{X}_{i\bullet\bullet}$	average of all results day i
$\overline{X}_{\bullet\bullet\bullet}$	average of all results

6 Overview of the General Precision Evaluation Experiment

6.1 General Guidelines

Proper evaluation of an analytical device requires:

- sufficient time to become familiar with the mechanics of operation and maintenance of the device according to the manufacturer's instructions;
- sufficient time to become familiar with the steps of the evaluation protocol;
- maintenance of the device in proper quality control during the entire period of the evaluation;
- sufficient data and an appropriate experiment that is of sufficient duration to generate adequate samples. (Data and experiment duration are critical in that precision estimates should have a sufficient number of degrees of freedom. This should properly reflect long-term performance of the device during a period of routine workload in the laboratory or for the customers); and
- statistically valid data analysis procedures.

How "sufficient data" is defined will depend on the ultimate use for the data and how well the precision of the device is determined.

6.2 Device Familiarization Period

The first step is to become familiar with the device, and with all aspects of set-up, operation, maintenance, and other factors in its routine laboratory use. Users can do this after or concurrently with the training period suggested by the manufacturer.

6.3 Protocol Familiarization Period

The first five operating days of the precision evaluation experiment should be used to become familiar with the experimental protocol itself. While practicing the experiment, any serious problems with the device should be detected. Data should be collected during this period, because it may be useful at the end of the experiment. Preliminary acceptability tests can be performed for precision and other characteristics not addressed in this guideline (such as linearity and drift) as described in NCCLS document EP10—*Preliminary Evaluation of Quantitative Clinical Laboratory Methods*.

6.4 Precision Evaluation Experiment

Once familiarity with the device is achieved, the precision evaluation experiment can be started. A minimum of 20 operating days is recommended for the precision evaluation experiment. Because day-to-day imprecision may be large, performance must be evaluated long enough to ensure that the separate components of error are adequately covered by the experiment.

During each of the testing days, two separate runs (when "runs" are an important component of the target device operating procedure) with two test samples at each of at least two levels of analyte concentration should be analyzed. In addition to the test samples, at least one quality control sample should be analyzed in each run. The laboratory's routine quality control procedures and materials (if appropriate) should be used during the evaluation.

If "runs" do not constitute an aspect of the device under consideration, then the four samples at each level should be analyzed as two sets of pairs, under repeatability conditions, at different times during the day. The paired results should be treated the same as two results obtained in the same run.

If a manufacturer of IVD devices or a method developer wishes to obtain estimates of the reproducibility of the method, he/she should run the protocol on multiple devices, in different locations, using different operators on each device. This is described more fully in Section 11.

6.5 Completing the Precision Experiment

After this protocol familiarization period, the experiment should be continued for 15 (or more) days. At the end of each five operating days, the control limits should be recalculated on a set of quality control charts and all data checked for acceptability. If this process identifies outliers, every attempt should be made to determine the cause of the problem. Data may not be rejected without valid justification, since this will lead to understatement of precision performance. When the precision experiment is completed, the appropriate statistical calculations on the data are performed. If the evaluator determines that a "learning curve" (or trend) is apparent during the protocol familiarization period, the earliest data may be excluded and replaced with an equal amount of data collected at the end of the originally scheduled evaluation period.

6.6 Comparison to Other Precision Evaluations

Other procedures sometimes used for evaluating precision consist of a single run of 20 observations for repeatability, or taking a single observation (or just a few) at a given concentration each day for 10 or 20 days for total imprecision (usually incorrectly calculated and erroneously labeled day-to-day precision). These procedures have serious drawbacks, since they fail to include significant sources of variation, and are specifically not recommended in this protocol.

When a single run is used to estimate repeatability (within-run imprecision), there is a significant risk that the operating conditions in effect at the time of that single run may not reflect usual operating parameters, thus adversely affecting the estimate. Furthermore, there is no way to determine how representative of expected performance that single run may be. For this reason, this document recommends that repeatability be estimated by "pooling" the within-run precision estimates over many runs, thus ensuring a more robust and representative estimate that should represent future performance under a variety of routine conditions.

Intermediate measures of precision calculated with procedures in this document will be independent of the number of days and runs within a day used to estimate it (which traditional methods are not). These procedures correctly combine the effect of repeatability, and between-run and between-day components of precision (which will vary in relative size from method to method), and avoid the error of using incorrect terms for precision (such as "day-to-day").

Manufacturers, regulators, and purchasers of commercial methods should ensure that an appropriate design has been used to account for all major sources of variation, and that proper statistical methods have been used to correctly adjust for number of observations per run and per day.

7 Statistical Power of Precision Estimates

7.1 Precision and Confidence

When designing an evaluation experiment, it must be decided beforehand how well the actual precision of the device is to be determined. Each time a certain precision protocol is run, an estimate of the precision

of the device is obtained. When this same protocol is rerun in the same laboratory with a device that is in control, a different estimate of the precision will result even though the actual precision is the same.

These estimates of precision might be expected to scatter around the "true" value, and the estimates obtained from more observations to cluster more closely around the "true" precision. In general, a larger number of observations leads to more confidence in an estimate, and the more confidence in an estimate, the more one has in the "statistical power" to detect performance that is different from the claim.

7.2 Statistical Comparison with the Manufacturer

With this precision evaluation experiment, estimates can be compared for repeatability and other measures of precision with those from the manufacturer. The statistical power of such comparisons can be calculated, that is, how much the estimates statistically differ from claimed performance based on the number of degrees of freedom of these estimates.

This extremely important concept can be used to illustrate that an estimate of repeatability based on 100 degrees of freedom can detect relatively small deviations from claimed performance. Likewise, an estimate of repeatability based on only, for example, 10 degrees of freedom will detect only major departures from claimed performance and thus, a test based on such an estimate has low statistical power.

If the estimate has 40 degrees of freedom, there is a greater statistical power and the estimate can detect smaller, though still clinically important, departures from claimed performance. This is an important aspect in the design of any evaluation experiment.

8 Device Familiarization Period

8.1 Purpose

The operation, maintenance procedures, methods of sample preparation, and calibration and monitoring functions required must be learned. Many manufacturers of clinical chemistry devices provide operator training. The device should be set up and operated in the individual laboratory long enough to understand all of the procedures involved to avoid problems during the actual evaluation of its performance. This should include testing actual sample material, including pools, controls, leftover serum (if appropriate), or any other test materials appropriate for the device.

All possible contingencies (such as error flags, error correction, calibration, etc.) that might arise during routine operation should be carefully monitored. Data should not be collected during this period. The device familiarization period is not complete until the user can demonstrate that he/she can operate the device properly.

8.2 Duration

A five-day familiarization period is adequate for most devices. A shorter or longer period may be appropriate, depending on the complexity of the device and the skill level of the operator.

9 Protocol Familiarization Period (for Users and Manufacturers)

9.1 Purpose

An evaluation experiment often involves steps not ordinarily encountered during routine laboratory conditions. To keep these unfamiliar steps from adversely affecting the results of the evaluation experiment, the experiment should be practiced for some time before starting the protocol. Use of this

period will ensure understanding of the protocol. The experiment should be run as described in the next section using the regular test materials and quality control materials in the laboratory.

9.2 Duration

This protocol familiarization should be continued until data are obtained without operational difficulty for a minimum of five operating days. This period can be extended as necessary for complex devices.

9.3 Use of Data

The data collected without operational difficulty during those five or more days should be incorporated into the estimation of precision along with data collected during subsequent operation of the protocol, if in the opinion of the evaluator these data are consistent with subsequent data. All data should be subjected to quality control acceptability checks as described below.

9.4 Quality Control Procedures

It is assumed that the device is operating in a stable condition while collecting the data during the protocol familiarization stage. To justify this assumption, the performance of the device should be monitored with quality control samples and routine quality control procedures. The trial control limits should be calculated after completing this phase of data collection. If these trial control limits do not reasonably agree with the manufacturer's performance claims for the device, the manufacturer should be contacted before the experiment is continued.

9.5 Additional Evaluations

While practicing the experiment, other features of the device can be checked. Linearity, trueness (recovery), or any other feature not discussed in these guidelines can be tested. These tests should be used to see if there are any serious problems with the device. If there are any problems, the manufacturer should be contacted to determine the cause of the problem. The decision of whether the device is acceptable should not be made solely on the basis of these limited preliminary tests.

9.6 Preliminary Precision Evaluation

At or near the end of the protocol familiarization period, an initial evaluation of repeatability should be conducted. Twenty aliquots of an appropriate test material (or a complete "batch" if less than 20) should be assayed in sequence. Ideally, two or more concentration levels should be used. The standard deviation and coefficient of variation of the results should be calculated. If a considerable discrepancy from expected results is found, the manufacturer should be contacted and no further testing should be conducted until the problem is solved. It should be emphasized that this single run test *is not* sufficient to judge the acceptability of the device. It can only identify problems that should be solved before continuing the evaluation. These data are used only for this one-time verification.

10 Precision Evaluation Experiment

This describes the basic protocol for estimating repeatability and within-laboratory precision for a single device, or a method as used in a single laboratory. It is for use by laboratories for estimating the precision of their device or method. It is also the basic protocol for use by manufacturers and method developers, if they follow the additional guidelines described in Section 11.

10.1 Components of Precision

The main objective of the precision evaluation experiment is to estimate the precision of the device or measurement method as used on a single instrument in a single laboratory. Intuitively, this precision is the variability of the device when used over an indefinitely long period. To some degree, several sources of variability contribute to this long-term precision. Generally, it is sufficient to design the experiment so that all these sources will influence the within-laboratory precision estimate without trying to determine the relative size of each source or component. Terms used to describe the time-related components of precision include:

- repeatability;
- between-run precision;
- within-day precision;
- between-day precision; and
- within-laboratory precision.

Of these, the repeatability and within-laboratory precision are generally of most interest. The experiment described in this section was designed to provide estimates of within-laboratory precision and repeatability of the device during operation in the laboratory. It was not attempted to incorporate specifically in this experiment separate estimates of other possibly significant sources of variability such as calibrator or reagent lot differences or technologist/operator differences; however, in Section 11 it is suggested that manufacturers should include such factors, as well as variability between devices in different locations. Other factors that influence precision, such as sample preparation, test material stability, carryover, and drift (refer to NCCLS document EP10—*Preliminary Evaluation of Quantitative Clinical Laboratory Methods*) are included in this protocol as sources of within-laboratory imprecision, but are not estimated separately.

10.2 Reagents and Calibration Materials

A single lot of reagents and calibration materials may be used for the entire protocol, but interpretation (and explicit labeling, when appropriate) of results must include this fact, and results may underestimate true long-term, within-laboratory (or within-device) precision. Introducing several lots of these materials will increase the observed variability, and although the experiment does not allow for separately estimating the effects of these factors, it may better represent the real precision performance of the device.

10.3 Test Materials

10.3.1 Matrix

The test materials should be selected to simulate the characteristics of the appropriate clinical samples. Stable, frozen pools are preferred when appropriate and available. When necessary, stable, commercially available, protein-based materials may be used.

10.3.2 Concentrations

Test materials should be chosen carefully by considering several criteria. Two concentrations are recommended, although more may be used. In this protocol precision is estimated separately for each level tested; there is no pooling or averaging across levels. If the precision estimates or relative precision

estimates are the same at these levels, then there is evidence of constant precision (or relative precision). If the estimates are not similar, then there may be a need to test more levels.

Concentrations that span a significant portion of the measuring range of the device should be selected whenever possible. If more than two concentrations are available, additional concentrations as close as possible to the "medical decision levels" used in the laboratory should be chosen. To compare evaluation results to published performance claims, concentrations should be chosen that correspond to the levels in those claims.

When establishing claims, a high level, a low level, and a level near a decision point should be tested. If the three levels show constant precision or constant relative precision, then three levels are sufficient. If the estimates are not similar or if there are large differences in the precision estimates at the three levels, then more levels should be tested to fully describe the performance of the method.

10.4 Number of Runs and Days

10.4.1 General Guidelines

The experiment and calculations described in this document are one example of an evaluation design. This experiment and its calculations are an example of a balanced design (a fully nested Model II ANOVA), which is appropriate for most clinical chemistry systems and devices. Other designs may be more appropriate for specific systems, but the required calculations and statistical interpretations will be different.

The precision evaluation experiment requires a sufficient amount of data, so the estimates of precision properly reflect the true precision parameters of the device. A minimum of 20 acceptable operating days is generally necessary to achieve this, except in situations where this is known not to be a factor. During the first five days of the experiment, the user should become familiar with the protocol as described in Section 9.

A short-run method has a run duration of less than two hours, while a long-run method (such as RIA) has a considerably longer "run," generally done once per shift. For long-run methods, the one run per day procedure in Appendix C should be used; for short-run procedures, the test samples may be tested anywhere in the run. For the purposes of the analysis of variance, an evaluation run is a discrete time period of data collection designed to enable the estimation of variability (or drift) within a day. For some devices, such as random-access, discrete, or unitary devices, the concept of a "run" may not be appropriate. In this case, samples should be run in pairs under repeatability conditions at random times throughout a working shift to simulate the actual operation of the device.

10.4.2 Specific Procedures

See Sections 6.2 (Device Familiarization Period) and 6.3 (Protocol Familiarization Period) for initial steps in the evaluation process.

The following steps shall be taken within each day:

(1) Analyze two runs or batches.

(2) If a run must be rejected because of quality control procedures or operating difficulties, conduct an additional run *after* an investigation is conducted to identify and correct the cause of the problem.

(3) Within each run or batch, analyze two aliquots of test material for each concentration used.

(4) Include in each run the quality control samples ordinarily used to judge the acceptability of the run or day.

(5) Change the order of analysis of test materials and quality control samples for each run or day.

(6) To simulate actual operation, include at least ten patient samples in each run whenever possible.

(7) Separate the runs performed each day by a minimum of two hours.

10.5 Recording the Data

Appendix A contains examples of data recording sheets to summarize data. This type of summary is valuable in the statistical analysis described below. If the number of runs, days, or observations is changed, a similar sheet should be created, the resulting data transcribed onto it, and the necessary calculations adjusted accordingly.

10.6 Quality Control Procedures

10.6.1 General Guidelines

Normal quality control procedures should be conducted during the precision evaluation experiment. At least one quality control sample at an appropriate concentration should be included in each run. If two or more concentrations for quality control are ordinarily used, this method should be continued throughout the evaluation experiment. For guidance on quality control practices refer to NCCLS document C24—*Statistical Quality Control for Quantitative Measurements: Principles and Definitions*.

10.6.2 Statistical Quality Control Charts

Preliminary statistical quality control charts should be set up for the device at the end of the protocol familiarization period (i.e., the first five acceptable days of the precision data collection period). The following procedure should be followed:

(1) Calculate the center lines, warning limits, and out-of-control limits from these initial data according to usual practices.

(2) Plot all subsequent quality control data on the charts.

(3) If at any point an out-of-control condition is detected, determine the cause, eliminate the offending point, and then repeat the run. It is suggested, since there is low statistical power with these preliminary estimates, that ±3 SDs be used as indications for investigation, and ±4 SDs be used for rejection. It is not acceptable to simply rerun a control sample to see if the new point is inside the control limit.

(4) After each of the five days of data collection, recalculate the center lines and control limits of each chart from all acceptable data collected thus far.

(5) If the previously acceptable results are now unacceptable, continue the precision experiment to obtain the proper number of days.

(6) Maintain a record of the number of rejected runs.

10.7 Detection of Outliers

A detection criterion for outliers must be defined to use during the precision evaluation experiment. The detection criterion is needed to be certain that operational problems will not unduly distort the resulting data and precision estimates.

Assuming appropriate quality control procedures will be used during the experiment, a fairly weak (low power) test is suggested to detect gross outliers in the data. The outlier test is derived from the data collected during the preliminary precision test. Data collected during each run of the precision evaluation experiment are in pairs (duplicates). The following test should be used:

(1) If the absolute value of the difference between the replicates exceeds 5.5 times the standard deviation determined in the preliminary precision test (see Section 9.6), the pair should be rejected.

(2) If such an outlier is found, the cause of the problem should be investigated, and the run repeated for that analyte. The value 5.5 is derived from the upper 99.9% value of the normalized range for the difference between two observations. **NOTE:** This test should be used when the concentration of the preliminary test material is reasonably close to the concentration of the evaluation test material.

The evaluator may wish to schedule additional days of evaluation at the outset of the investigation, to allow for potential run rejections, if needed. 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 a valid variability assessment.

10.8 Statistical Calculations for Precision

After collecting the data and transcribing them onto an appropriate recording sheet, the calculations described in this section should be performed. A sample completed recording sheet can be found in Appendix B, along with the associated calculations. Separate calculations should be performed for each concentration, and all data checked against the outlier criterion described in Section 10.7.

10.8.1 Repeatability Estimate

The estimate of repeatability is derived from the following formula:

$$S_r = \sqrt{\frac{\sum_{i=1}^{I} \sum_{j=1}^{2} (X_{ij_1} - X_{ij_2})^2}{4I}} \qquad (1)$$

where:

I	=	total number of days (generally 20)
j	=	run number within-day (1 or 2)
X_{ij_1}	=	result for replicate 1, run j on day i
X_{ij_2}	=	result for replicate 2, run j on day i.

Two results are needed on each of two runs for every day to use the above formula. If only one run is available on a given day, this formula can still be used (except j=1). As long as there are no more than 10% of the evaluation days with missing runs (i.e., only one run) in the two-run-per-day experiment, the resulting statistical calculations will be valid. See Appendix C for formulas to use when there is only one run on each day.

10.8.2 Estimates of Within-Device (or Within-Laboratory) Precision

Several quantities are required to determine estimates of precision for a device or within a laboratory. The calculations below will be needed:

$$A = \sqrt{\frac{\sum_{i=1}^{I} (\overline{X}_{i_1\bullet} - \overline{X}_{i_2\bullet})^2}{2I}} \qquad (2)$$

where:

I = number of days (with two runs)
$X_{i_1\bullet}$ = average result run 1, day i (average of the two replicates)
$X_{i_2\bullet}$ = average result run 2, day i (average of the two replicates).

A is calculated by squaring the difference between the first run analysis average and the second run analysis average for each day, summing up these quantities for all days, dividing by 2I, and taking the square root. This calculation should not be used if the data were generated on a day with only one run.

The second quantity is:

$$B = \sqrt{\frac{\sum_{i=1}^{I} (\overline{X}_{i\bullet\bullet} - \overline{X}_{\bullet\bullet\bullet})^2}{I - 1}} \qquad (3)$$

where:

I = number of days
$X_{i\bullet\bullet}$ = average of all results day i
$X_{\bullet\bullet\bullet}$ = average of all results.

This is the standard deviation of the daily means, as identified in Data Sheet #3 in Appendix A.

The following are then calculated:

$$S_{dd}^2 = B^2 - \frac{A^2}{2}$$

where:

S_{dd} = estimate of between-day standard deviation.

$$S_{rr}^2 = A^2 - \frac{S_r^2}{2}$$

where:

S_{rr} = estimate of between-run standard deviation.

(set to 0 if negative).

Setting the (possibly) negative variance components to zero follows a widely used convention in statistics. If these calculations are performed with a software package, adherence to the above convention should be assured.

The estimate of within-device or within-laboratory precision is then calculated with the following standard deviation formula:

$$S_T = \sqrt{S_{dd}^2 + S_{rr}^2 + S_r^2} \quad (4)$$

A different result will be obtained from this formula for S_T compared to that obtained by calculating the standard deviation of all data observed (without regard to day or run). The above formula is the correct way to estimate the precision of a device, because it properly weights the repeatability as well as the between-day and between-run components. The coefficient of variation corresponding to this estimate of precision should be calculated by dividing S_T by the concentration of the test material and multiplying by 100. The result should be expressed as a percentage.

10.9 Comparison with Manufacturers' Claims or Other Performance Criteria

The precision estimates obtained in the previous section should be compared to performance claims for the precision of the device. The chi-square **(χ^2)** statistic as described below should be used. To use this method, the performance claim is expressed as a point estimate (i.e., a standard deviation). The repeatability and intermediate precision estimates should be compared separately.

10.9.1 Repeatability Comparison

The performance claim standard deviation (σ_r) should be denoted. The chi-square test uses the square of both the user's and manufacturer's estimates of repeatability. The number of degrees of freedom associated with S_r^2 (the user's estimated within-run variance) must be known. In the experiment described in this protocol, S_r^2 will have as many degrees of freedom as there were data pairs (replicates within runs) used to calculate it. Thus, this will be equal to the number of runs during the experiment, which will be denoted R below. The test involves calculating the following:

$$\chi^2 = \frac{S_r^2 \bullet R}{\sigma_r^2} \quad (5)$$

where:

S_r^2 = square of the user's estimated repeatability variance

σ_r^2 = square of the manufacturer's claim of repeatability variance

R = the total number of runs (degrees of freedom for S_r^2).

The calculated χ^2 should be compared with a statistical table of χ^2 values, using the upper 95% critical value with R degrees of freedom (see Table 1). If the calculated value is less than this table value, then the estimate is not significantly different from the claimed value, and this part of the precision claim is accepted.

NOTE: The estimate may be larger than the manufacturer's claim, and still *not* be significantly different.

Table 1. Critical Values of Chi-Square

df of User Variance Estimate	95% Critical Value	99% Critical Value
5	11.1	1.51
6	12.6	16.8
7	14.1	18.5
8	15.5	20.1
9	16.9	21.7
10	18.3	23.2
11	19.7	24.7
12	21.0	26.2
13	22.4	27.7
14	23.7	29.1
15	25.0	30.6
16	26.3	32.0
17	27.6	33.4
18	28.9	34.8
19	30.1	36.2
20	31.4	37.6
25	37.7	44.3
30	43.8	50.9
35	49.8	57.3
40	55.8	63.7
50	67.5	76.2
60	79.0	88.4
70	90.5	100.4
75	96.2	106.4
79	100.7	111.1
80	101.9	112.3
90	113.1	124.1
100	124.3	135.6

10.9.2 Comparison of Precision Estimate

A chi-square test similar to that described above should be used to compare the estimate of within-laboratory (within-device) precision to that claimed by the manufacturer, or to that required by the medical application at the user's institution. Unlike the repeatability estimate, however, computing the exact number of degrees of freedom for S_T involves a complicated calculation. Because of the structure of the protocol, the user cannot assume that all observations are independent, a necessary assumption before the customary estimate for degrees of freedom (total number observations minus one) can be used. The formula below for T degrees of freedom for S_T takes into account this lack of independence.

Let:

ME = S^2_r (mean square for within-run, or repeatability variance)

MR = $2A^2$ (mean square for runs)

MD = $4B^2$ (mean square for days)

where:

S_r is defined in Section 10.8.1, and A and B are defined in Section 10.8.2.

$$\text{Then, } T = \frac{I(2ME + MR + MD)^2}{2ME^2 + MR^2 + \frac{I}{I-1}MD^2} \quad (6)$$

The nearest integer to this calculated value should be used as the appropriate degrees of freedom for S_T.

$$\text{Using this value, the appropriate statistic is as follows: } \chi^2 = \frac{S_T^2 \times T}{\sigma_T^2} \quad (7)$$

where:

S_T^2 = square of the user's estimate of within-laboratory (within-device) standard deviation

σ_T^2 = square of manufacturer's claim of device standard deviation, or medically required standard deviation

T = degrees of freedom for S_T.

If the calculated χ^2 is less than the critical upper 95% χ^2 value (from Table 1), the precision performance is assumed to be acceptable.

If the calculated χ^2 is greater than the critical upper 95% χ^2 value, the precision performance is not within the claimed limits, or is not acceptable for the defined medical application.

The user's estimate can be larger than the SD claimed by the manufacturer and still be acceptable. Since the user experiment is based on a limited number of observations, there are expected sampling errors of the calculated S_r and S_T around the true values. The larger the user experiment, the closer the estimates will be to the true value. The chi-square test is used to determine if the user's estimates are *significantly* larger than those provided by the manufacturer.

Including a list of analyte-specific, "acceptable" standard deviations is not within the scope of this document. It is suggested that the medical staff of the user's institution be consulted or the technical literature examined to develop an appropriate numerical definition or standard for acceptable standard deviation of each analyte.

11 Use of These Guidelines by Manufacturers to Establish Precision Performance

11.1 Factors to be Considered

The experiment described in this document can be used by manufacturers to establish precision performance claims for repeatability, device precision, and related coefficients of variation. However, the goal of the manufacturer should be to establish these point estimates with sufficient rigor, so they will be valid over the wide variety of operating environments that individual users may encounter in the routine use of the method, device, or instrument.

The manufacturer may choose to employ a single reagent lot, calibration cycle, device, and operator for a minimum of 20 days to estimate the device's precision. This approach minimizes the effects of factors (which increase long-term imprecision) and increases the manufacturer's risk that individual users may

not be able to achieve similar results in their laboratories. This risk may be reduced by incorporating multiple locations, devices, operators, reagent lots, calibrator lots, and calibration cycles (if appropriate), which will generally increase the precision standard deviation. Including additional sources of variation should better reflect the range of results that will be experienced by customers.

If these experiments are used by manufacturers to establish precision claims, the resulting labeling *must* include a statement regarding the number of locations, days, runs, devices, operators, calibration cycles, calibrator lots, and assay reagent lots that were included in the evaluation.

11.2 Incorporating Multiple Factors

Two approaches are available for incorporating the effects of multiple factors in the data. The first method is to perform the basic two-runs-per-day experiment described in this document, but using multiple reagent lots, calibration cycles, operators, and instruments over the course of the 20 or more days of the evaluation experiment. The data may be analyzed and summarized according to the formulae provided, but will now reflect the influence on precision performance of those factors which were incorporated into the design of the experiment. The estimates will then better reflect the range of precision likely to be experienced in users' laboratories. If this approach is used, the manufacturer should consider extending the protocol to more than 20 days.

The second method permits the use of multiple instruments providing more than two runs per day. In this situation, the general nested analysis of variance should be used to determine the components of variance that pertain to each individual instrument, incorporating multiple reagent and calibrator lots, calibration cycles, and operators. The estimates for each instrument may then be "pooled" to create the precision performance claims, or results may be presented individually. This pooling may be done *only* when *each* instrument is evaluated with multiple operators, reagent lots, and calibrator lots, *not* when there is only one such factor level per device. This method will reflect variations in precision performance between different instruments without incorporating the actual instrument-to-instrument component which would not be applicable to single-instrument users. When this procedure is used, an additional intermediate precision component will be produced for between-device precision.

Reproducibility precision can be estimated by conducting precision experiments in different locations[7] (for example, multicenter trials), and isolating between-laboratory variance from the within-laboratory components. Variability between laboratories is often the largest single component of variability. Therefore, reproducibility is often the largest component of method variability, and repeatability is often the smallest component—the irreducible inherent variability of the method. Intermediate components (between run, between day, etc.) are often sources under the watch of the individual laboratory. The exact calculations for multilaboratory or multi-instrument designs are beyond the scope of this document but may be found in standard references on the analysis of variance or in ISO 5725-2.[7]

It must be noted that regulatory agencies may require identification, evaluation, and estimation of variance components beyond those employed in a user precision evaluation when this guideline is employed by manufacturers. In cases where additional components need evaluation, it is suggested that a statistician be consulted for appropriate experimental designs.

11.3 Format for Statement of Claims

Labeled claims for precision performance must include the following information, except where noted as optional:

- concentrations at which claim is made;

- point estimates (single value parameter estimate) of repeatability standard deviation at every level tested;
- repeatability percent coefficient of variation (optional);
- point estimate of within-device (or within-laboratory) precision standard deviation at every level tested;
- within-device or within-laboratory precision percent coefficient of variation (optional);
- confidence intervals on repeatability and device standard deviation (optional);
- actual number of days involved in the experiment, and number of sites;
- actual total number of runs (if applicable);
- total number of observations (optional);
- number of instruments/devices used in the evaluation, and how results were pooled;
- number of reagent lots; and
- number of calibration cycles and calibration lots.

A manufacturer may elect to include a table of expected maximum observed standard deviation (tolerance limit) for the repeatability and within-laboratory precision SDs, indexed by degrees of freedom (df). This will provide the user of the method with a benchmark to indicate that small verification experiments may result in calculated estimates slightly higher than the published SD point estimate and still demonstrate statistically equivalent precision. Table 2 may be used for the multipliers of the claimed SDs to create a table that may look like this:

Repeatability SD published: 10.5 at 40 mg/dL

df for User Experiment	Acceptable SD Maximum
10	14.2
20	13.1
30	12.6
40	12.4
100	11.7

The purpose of such a table in a labeling claim would be to simply illustrate the sometimes confusing fact that a user verification estimate can sometimes be higher than the published SD point estimate, and still be used to verify a claim.

Table 2. Tolerance Factors for User SD Estimates

df for User SD Estimates	Upper 95% Tolerance Limit for 95% of User Estimates[*†]
10	1.35
20	1.25
30	1.20
40	1.18
50	1.16
60	1.15
70	1.14
80	1.13
90	1.12
100	1.11

[*] Multiply point estimate from manufacturer experiment by this factor to obtain the upper tolerance limit.
[†] From Hald A. *Statistical Theory with Engineering Applications*. New York: Wiley; 1952:277.

References

[1] ISO. *Statistics – Vocabulary and symbols – Part 1: Probability and General Statistical Terms*. ISO 3534-1. Geneva: International Organization for Standardization; 1993.

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

[3] ISO. In vitro *diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to catalytic concentration of enzymes in calibrators and control materials*. ISO 18153. Geneva: International Organization for Standardization; 2003.

[4] ISO. *International Vocabulary of Basic and General Terms in Metrology*. Geneva: International Organization for Standardization; 1993.

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

[6] 42 CFR Part 493. *Medicare, Medicaid, and CLIA Programs; Laboratory Requirements Relating to Quality Systems and Certain Personnel Qualification*; February 28 1992.

[7] 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: International Organization for Standardization; 1994.

Additional References

Bauer S, Kennedy JW. Applied statistics for the clinical laboratory. *J Clin Lab Auto.* I-XI (a series of 11 individual articles). 1982; 1983; 1984.

Box GEP, Hunter WG, Hunter JS. *Statistics for Experimenters. Study of Variation.* New York: John Wiley and Sons; 1978.

CAP. Proceedings of the 1976 Aspen Conference, Analytical goals in clinical chemistry. Skokie, IL: College of American Pathologists; 1977.

Appendix A. Sample Data Recording Sheets

Data Sheet #1: Precision Evaluation Experiment

Concentration:
Analyte:
Device:

Operator:
Reagent Source/Lot:
Calibrator Source/Lot:

		Run 1			Run 2			
Day #	**Date**	**Result 1**	**Result 2**	**Mean**	**Result 1**	**Result 2**	**Mean**	**Daily Mean**

Appendix A. (Continued)

Data Sheet #2: Precision Evaluation Experiment

Analyte/Concentration: Device:

	Run 1	Run 2	
Day #	$(\text{Rep 1} - \text{Rep 2})^2$	$(\text{Rep 1} - \text{Rep 2})^2$	$(\text{Mean Run 1} - \text{Mean Run 2})^2$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
Sums	(1)	(2)	(3)

Appendix A. (Continued)

Data Sheet #3: Using Results Labeled (1), (2), and (3) from Data Sheet #2

Section

10.8.1 $S_r = \sqrt{\frac{(1) + (2)}{4I}} =$ _____

where I = number of days

10.8.2 $A = \sqrt{\frac{(3)}{2I}} =$ _____

from Data Sheet #1

10.8.2 B = Standard deviation of "Daily Means" = _____

10.8.2 $S_T = \sqrt{\frac{2B^2 + A^2 + S_r^2}{2}} =$ _____

Calculation of T (degrees of freedom for total standard deviation estimate)

$ME = S_r^2 =$ _____ $MR = 2A^2 =$ _____ $MD = 4B^2 =$ _____

$$T = \frac{I(2ME + MR + MD)^2}{2ME^2 + MR^2 + \frac{I}{I-1} MD^2}$$

= _____

= _____(rounded to nearest integer)

Appendix A. (Continued)

Data Sheet #4: Precision Evaluation Experiment Comparison to Claims

Repeatability	
	User Concentration Level =
User SD*	Claim Concentration Level =
User Variance (SD^2)	Degrees of Freedom (R)
Performance Claim SD	
Variance (SD^2)	
(I) (User Variance ÷ Claim Variance) • R =	
(II) Critical Chi-square (from Table 1)	
☐ Claim Rejected (I > II) ☐ Claim Accepted (I ≤ II)	

Device/Method Precision	
	User Concentration Level =
User SD	Claim Concentration Level =
User Variance (SD^2)	Degrees of Freedom (T)
Performance Claim SD	
Variance (SD^2)	
(I) (User Variance ÷ Claim Variance) • T =	
(II) Critical Chi-square (from Table 1)	
☐ Claim Rejected (I > II) ☐ Claim Accepted (I ≤ II)	

* "SD" means standard deviation, and refers to the S_r (calculated), S_T (calculated), or manufacturer claim.

Appendix B. Example of Completed Sample Data Recording Sheets

Data Sheet #1: Precision Evaluation Experiment

Concentration: High
Analyte: Glucose
Device: XYZ

Operator:
Reagent Source/Lot: AA—Lot 87011
Calibrator Source/Lot: AA—Lot 87011

		Run 1			Run 2			
Day #	**Date**	**Result 1**	**Result 2**	**Mean**	**Result 1**	**Result 2**	**Mean**	**Daily Mean**
1	7/8	242	246	244	245	246	245.5	244.75
2	7/9	243	242	242.5	238	238	238	240.25
3	7/10	247	239	243	241	240	240.5	241.75
4	7/11	249	241	245	250	245	247.5	246.25
5	7/14	246	242	244	243	240	241.5	242.75
6	7/15	244	245	244.5	251	247	249	246.75
7	7/16	241	246	243.5	245	247	246	244.75
8	7/17	245	245	245	243	245	244	244.5
9	7/18	243	239	241	244	245	244.5	242.75
10	7/21	244	246	245	247	239	243	244
11	7/22	252	251	251.5	247	241	244	247.75
12	7/23	249	248	248.5	251	246	248.5	248.5
13	7/24	242	240	241	251	245	248	244.5
14	7/25	246	249	247.5	248	240	244	245.75
15	7/28	247	248	247.5	245	246	245.5	246.5
16	7/29	240	238	239	239	242	240.5	239.75
17	7/30	241	244	242.5	245	248	246.5	244.5
18	7/31	244	244	244	237	242	239.5	241.75
19	8/1	241	239	240	247	245	246	243
20	8/4	247	240	243.5	245	242	243.5	243.5

Appendix B. (Continued)

Data Sheet #2: Precision Evaluation Experiment

Analyte/Concentration: Glucose/High Device: XYZ

	Run 1	Run 2	
Day #	$(\text{Rep 1} - \text{Rep 2})^2$	$(\text{Rep 1} - \text{Rep 2})^2$	$(\text{Mean Run 1} - \text{Mean Run 2})^2$
1	16	1	2.25
2	1	0	20.25
3	64	1	6.25
4	64	25	6.25
5	16	9	6.25
6	1	16	20.25
7	25	4	6.25
8	0	4	1.00
9	16	1	12.25
10	4	64	4.00
11	1	36	56.25
12	1	25	0
13	4	36	49.00
14	9	64	12.25
15	1	1	4.00
16	4	9	2.25
17	9	9	16.00
18	0	25	20.25
19	4	4	36.00
20	49	9	0.00
Sums	(1) 289	(2) 343	(3) 281.00

Appendix B. (Continued)

Data Sheet #3: Using Results Labeled (1), (2), and (3) from Data Sheet #2

Section

10.8.1 $$S_r = \sqrt{\frac{(1) + (2)}{4I}} = 2.81 \quad \text{(B1)}$$

where I = number of days

10.8.2 $$A = \sqrt{\frac{(3)}{2I}} = 2.65$$

from Data Sheet #1

10.8.2 B = Standard deviation of "Daily Means" = 2.34

10.8.2 $$S_T = \sqrt{\frac{2B^2 + A^2 + S_r^2}{2}} = 3.60$$

Calculation of T (degrees of freedom for device standard deviation estimate)

$ME = S_r^2 = 7.90$ $MR = 2A^2 = 14.0450$ $MD = 4B^2 = 21.9024$ (B2)

$$T = \frac{I(2ME + MR + MD)^2}{2ME^2 + MR^2 + \frac{I}{I-1} MD^2}$$

= 64.76

= 65 (rounded to nearest integer)

Appendix B. (Continued)

Data Sheet #4: Precision Evaluation Experiment Comparison to Claims

Repeatability	
	User Concentration Level = Gluc/High
User SD* 2.81	Claim Concentration Level = 240 mg/dL
User Variance (SD^2) 7.90	Degrees of Freedom (R) 40
Performance Claim SD 2.5	
Variance (SD^2) 6.25	
(I) (User Variance ÷ Claim Variance) • R = 50.56	
(II) Critical Chi-square (from Table 1) 55.8	
☐ Claim Rejected (I > II) ☑ Claim Accepted (I ≤ II)	

Device/Method Precision	
	User Concentration Level = Gluc/High
User SD 3.60	Claim Concentration Level = 240 mg/dL
User Variance (SD^2) 12.96	Degrees of Freedom (T) 65
Performance Claim SD 3.4	
Variance (SD^2) 11.56	
(I) (User Variance ÷ Claim Variance) • T = 72.65 (**NOTE:** Answers may vary slightly due to rounding of intermediate results.)	
(II) Critical Chi-square (from Table 1) 84.8	
☐ Claim Rejected (I > II) ☑ Claim Accepted (I ≤ II)	

* "SD" means standard deviation, and refers to the S_r (calculated), S_T (calculated), or manufacturer claim.

Appendix C. Additional Statistical Considerations

C1 Modifications for One Run per Day

For some devices, only one run per day may be needed. Correct and useful estimates of repeatability and within laboratory precision standard deviations for the device can still be obtained. However, the separation of precision into between-day and between-run, within-day components is not possible. The estimate of repeatability standard deviation should be calculated from the following formula:

$$S_r = \sqrt{\frac{\sum_{i=1}^{I} (X_{i1} - X_{i2})^2}{2I}} \qquad \text{(C1)}$$

where:

I = total number of days (generally 20)
X_{i1} = result for replicate 1 on day i
X_{i2} = result for replicate 2 on day i.

The procedures and specifics of the general protocol as described in the main document should be followed except for running only one run per day instead of two. **NOTE:** There are only half as many degrees of freedom in this estimate as there are with two runs per day.

C1.1 Increasing Degrees of Freedom

Two methods may be used to modify the protocol to increase the number of degrees of freedom for the repeatability estimate.

C1.1.1 Increase Length of Experiment

The number of days in the experiment may be increased, continuing to run only two aliquots of precision test material per run. The formula above may still be used for calculations. A minimum of 30 days is recommended.

C1.1.2 Increase Number of Aliquots

More than two aliquots of material within each run for the 20 days may be analyzed. If this method is used, the within-run standard deviation should be calculated from the following formula:

$$S_r = \sqrt{\frac{\sum_{i=1}^{I} \sum_{j=1}^{N} (X_{ij} - \overline{X}_{i\bullet})^2}{I(N-1)}} \qquad \text{(C2)}$$

where:

I = total number of days
N = number of replicate analyses per run
X_{ij} = result on replicate j in run on day i
$\overline{X}_{i\bullet}$ = average (mean) of all replicates on day i.

Appendix C. (Continued)

The number of degrees of freedom in this estimate is then l times the number of replicates per run minus 1 [l • (N - 1)]. Each run must contain the *same* number of replicates for this formula to be appropriate. **NOTE:** Factor 2 does not appear in the denominator, as now this formula uses the sum of squared deviations from the run mean, as opposed to the convenient shortcut of duplicate observation differences used in previous formulas (appropriate for only two observations).

C1.2 Within-Device or Within-Laboratory Precision Standard Deviation

The within device or within laboratory precision standard deviation estimates should be calculated with B from the following formulas. With only one run per day, the procedure differs somewhat from the formula described in the main protocol.

Calculate:

$$B = \sqrt{\frac{\sum_{i=1}^{I} (\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet})^2}{I - 1}} \qquad \text{(C3)}$$

where:

I = number of days
$\overline{X}_{i\bullet}$ = average replicates on day i
$\overline{X}_{\bullet\bullet}$ = average of all results over all days.

C1.2.1 Standard Error

B is the standard deviation of the daily means (generally called the *standard error* of the daily means). When only one run per day is performed, the estimate combines the between-day and between-run components of precision. This formula should be used regardless of the number of days or the number of replicates.

C1.2.2 Device or Laboratory Standard Deviation

The estimate of the device or laboratory precision standard deviation from the quantity B calculated above, and the within-run standard deviation estimate S_r, are as follows:

$$S_T = \sqrt{B^2 + \frac{N - 1}{N} S_r^2} \qquad \text{(C4)}$$

where:

N = number of replicates per run
B = standard deviation of daily means
S_r^2 = repeatability variance estimate (standard deviation squared).

This formula can be used regardless of which method is used to increase the number of observations for repeatability (additional days or additional replicates per run).

Appendix C. (Continued)

C1.2.3 Satterthwaite's Equation

Use Satterthwaite's equation to calculate the proper number of degrees of freedom for S_T. This is the only way to obtain the proper value for use in the chi-square test of claims described in Section 10.9.2. Use the following procedure:

$ME = S_r^2$ (mean square for within run)

$MD = N \times B^2$ (mean square for both runs and days)

Then, calculate T as:

$$T = \frac{I \times \left[(N-1) \times ME + MD\right]^2}{(N-1) \times ME^2 + \frac{I}{I-1} \times MD^2} \quad \text{(C5)}$$

Use the nearest integer to this quantity as the appropriate degrees of freedom for S_T.

C2 Other Estimates Available and Derivation of Formulas

Terminology that describes day-to-day or within-day precision has created confusion. Often "day-to-day" is erroneously used to mean precision over a long period of time. Additional confusion results because the parameters of the components of precision are independent of the type of experiment, while the calculations for these estimates differ greatly depending on the number of observations per run, runs per day, and number of days.

C2.1 Between-Day Precision

Statistically, day-to-day (more appropriately called "between-day") precision is the (adjusted) standard deviation of the daily means, after removing the effects of repeatability and between-run, within-day variability, on the daily averages. Think of it as an estimate of the variability of daily averages that you would expect if you could perform an infinite number of observations each day. If you conduct a single run each day, you can demonstrate that the variance of the daily averages has the following *expected value*:

$$\mathrm{Var}\left(\overline{X}\right) = \frac{\sum_{i=1}^{I}\left(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet}\right)}{I-1} = S_D^2 \quad \text{(C6)}$$

Expected value: $E\left(S_D^2\right) = \sigma_{dd}^2 + \frac{\sigma_r^2}{N}$

where:

$\overline{X}_{i\bullet}$ = average result on day I

$\overline{X}_{\bullet\bullet}$ = average of all results on all days

I = total number of days

Appendix C. (Continued)

σ_{dd}^2 = true (adjusted) between-day variance

σ_r^2 = true repeatability (within-run) variance

N = number of replicates per run.

As the number of replicates per run increases, the closer the estimate will be to the true parameter (i.e., the repeatability will have a lesser influence on the estimate). The quantity called B cannot be used on the protocol to estimate the between-day precision. An adjustment must be made to this quantity for it to be useful. The adjustment/estimation procedure depends on the number of runs per day and the number of observations per run, but *not* on the number of days used in the protocol (except in the proper calculation of the original estimates).

C2.2 Two Runs Per Day

For two runs per day and two observations per run, as described in the main protocol, the quantities A and B from Section 10.8.2 should be used to derive the following additional estimates:

Between-day standard deviation:

$$S_{dd} = \sqrt{B^2 - \frac{A^2}{2}} \qquad \text{(C7)}$$

The quantity S_{dd} is the estimate of the "true" adjusted between-day standard deviation, σ_{dd}.

Between-run, within-day standard deviation:

$$S_{rr} = \sqrt{A^2 - \frac{S_r^2}{2}} \qquad \text{(C8)}$$

If working with a new device, it may be useful to calculate these estimates for a better picture of the factors influencing the observed precision.

C2.3 Single Run per Day

For a single run per day and two or more observations per run as described in this appendix, the procedure is somewhat different. The between-day and between-run components of precision cannot be separated. The quantity called B in this case measures the sum of these two components. The only thing to do is to remove the effect of within-run variability from the estimate by calculating the following (N is the number of replicates/run):

$$S_{dd} = \sqrt{B^2 - \frac{S_r^2}{N}} \qquad \text{(C9)}$$

NOTE: The interpretation of the quantity S_{dd} is now the sum of the between-day and between-run within-day effects.

Appendix C. (Continued)

Also, in some instances, the quantity under the radical above may be negative, which can occur if the between-day true component value is small. If this occurs, then the estimate S_{dd} should be set to 0 (zero). This same caution should be applied to the estimates calculated above for two runs per day.

Data Calculation Sheet #1: Precision Evaluation Experiment

One Run per Day

Analyte/Concentration: Glucose/High Device: XYZ

Day#	Run Variance (S^2_r)	Run Mean
1	8.0	244
2	0.5	242.5
3	32.0	243
4	32.0	245
5	8.0	244
6	0.5	244.5
7	12.5	243.5
8	0.0	245
9	8.0	241
10	2.0	245
11	0.5	251.5
12	0.5	248.5
13	2.0	241
14	4.5	247.5
15	0.5	247.5
16	2.0	239
17	4.5	242.5
18	0.0	244
19	2.0	240
20	24.5	243.5
		Grand Mean = 244.13

Total WR Variance (S^2_r) = 7.225	Variance of Daily Means $\left(B^2\right)$ = 8.88
Total WR SD (S_r) = 2.69	SD of Daily Means $\left(B\right)$ = 2.98

NOTE: Data used are from the first run of the main protocol example.

Appendix C. (Continued)

Data Calculation Sheet #2: Precision Evaluation Experiment

One Run per Day

Analyte/Concentration: Glucose/High Device: XYZ

Calculation of Device (or laboratory) Precision Standard Deviation:

S_r^2 (from Sheet #1): 7.225

(C10)

B^2 (from Sheet #1): 8.88

N (# replicates per run): 2

$$S_T = \sqrt{B^2 + \frac{N-1}{N} S_r^2} = \sqrt{8.88 + \frac{1}{2}(7.225)}$$

=3.53

Calculation of proper number of degrees of freedom for S_T:

(C11)

$I = \#\ of\ days = 20$

$ME = S_r^2 = 7.225$

$MD = N \bullet B^2 = 17.76$

$$T = \frac{\left[(N-1)ME + MD\right]^2}{\frac{(N-1)ME^2}{I} + \frac{MD^2}{I-1}}$$

$$= \frac{(7.225 + 17.76)^2}{\frac{(7.225)^2}{20} + \frac{(17.76)^2}{19}}$$

$$= \frac{624.25}{19.21}$$

$$= 32.49$$

Use 32 for T.

NCCLS consensus procedures include an appeals process that is described in detail in Section 8 of the Administrative Procedures. For further information, contact the Executive Offices or visit our website at www.nccls.org.

Summary of Consensus/Delegate Comments and Committee Responses

EP5-A2: *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition*

General

1. A glossary would be useful either as a "lead-in" section or with definitions clearly called out in the text.

- **A section on "Definitions" has been added as recommended.**

2. The protocols for evaluated precision in both EP5-A and EP5-A2 do not strictly apply to hematology testing, since it is not possible to have the same blood specimen tested for several times within a day for 20 days. We can have protocols to estimate repeatability (as defined in EP5-A2), but most of the time other variance estimates (day-to-day, laboratory-to-laboratory, etc) will be confounded.

- **This concern is addressed in the third sentence in the Scope, which states: "These procedures may not be appropriate for some quantitative methods for which adequate test materials do not exist."**

3. I would have preferred to see procedures for the estimation of measurements of uncertainty (ISO).

- **The committee agrees that MU (measurements of uncertainty) is an important concept and that EP5-A2 is closely related. This is addressed in the fifth paragraph of the Foreword, which states that the precision estimates from EP5-A2 are components of measurement uncertainty, but that GUM (ISO Guide to the Expression of Uncertainty in Measurement) estimates of MU that comply with ISO procedures are beyond the scope of the document. This is because laboratory-specific estimates of uncertainty may contain components of error other than precision, and may involve corrections for bias.**

4. The recommended number of days (20) should be given in terms of 'repeats for a certain factor' (or degrees of freedom) to make it more general to any deviation from the base protocol. In addition, a factor that is expected to introduce more variability should be repeated more times than some other that is fairly constant.

- **See response to comment number 6, below. See also Appendix C for discussion of modifications.**

5. Formulas for calculating precision apply to experiments conducted exactly as described in the protocol. These formulas are totally or partially useless and sometimes misleading when even slight deviations from the protocol (e.g., having three replicates instead of two) are implemented. Currently, there are faster and more elegant ways for obtaining estimates of variance components. I think discussions should be based on the use of commercial statistical software (similar to EP6-A), while these formulas can be put together in an appendix for reference.

- **The formulae for unbalanced designs are quite complex and beyond the scope of this document. To be fully general for all situations would reduce the usefulness for less statistically sophisticated users, as explained in the Foreword (third paragraph). While many statistical software packages correctly calculate variance components, not all of them do. Further, these software packages can be difficult to interpret and are best left to the attention of professional statisticians.**

Section 7, Statistical Power of Precision Estimates

6. Section 7 does not provide any 'concrete' guideline for sample size. The discussion is purely academic.

- **It is the intent of the committee to specify a minimum number and guidance for when a situation might require larger numbers. Goal-based experimental design is beyond the scope of the document, but might be considered in a future version.**

Section 7.2, Statistical Comparison with the Manufacturer

7. Make degrees of freedom consistent. Use either 100 or 40 for both.

- **Section 7.2 presents guidance that is consistent with the general protocol, which allows the user to choose the number of samples to test. In this document, 100 and 40 are presented as different ends of the range of recommended replicates, not as a choice of one or the other.**

Section 10.9.1, Repeatability Comparison

8. In the description of equation (5) "square of the" is not needed.

- **While the committee agrees that "square of the" is not needed, the committee's opinion is that the current text is less confusing, since S_r is defined as the estimate of repeatability, which is squared in the equation.**

NOTES

The Quality System Approach

NCCLS subscribes to a quality 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 approach is based on the model presented in the most current edition of NCCLS document HS1—*A Quality System Model for Health Care.* The quality system approach applies a core set of "quality system essentials (QSEs)," basic to any organization, to all operations in any healthcare service's path of workflow (i.e., 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 quality system essentials (QSEs) are:

Documents & Records	Equipment	Information Management	Process Improvement
Organization	Purchasing & Inventory	Occurrence Management	Service & Satisfaction
Personnel	Process Control	Assessment	Facilities & Safety

EP5-A2 addresses the quality system essentials (QSEs) indicated by an "X." For a description of the other NCCLS documents listed in the grid, please refer to the Related NCCLS Publications section on the following page.

Documents & Records	Organization	Personnel	Equipment	Purchasing & Inventory	Process Control	Information Management	Occurrence Management	Assessment	Process Improvement	Service & Satisfaction	Facilities & Safety
					X C24 EP10 EP15 M29						M29

Adapted from NCCLS document HS1—*A Quality System Model for Health Care*.

Related NCCLS Publications*

C24-A2 **Statistical Quality Control for Quantitative Measurements: Principles and Definitions; Approved Guideline—Second Edition (1999).** This guideline provides definition of analytical intervals; plans for quality control procedures; and guidance for quality control applications.

EP10-A2 **Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition (2002).** This guideline addresses experimental design and data analysis for preliminary evaluation of the performance of an analytical method or device.

EP15-A **User Demonstration of Performance for Precision and Accuracy; Approved Guideline (2001).** This guideline demonstrates method precision and accuracy for laboratory analyte determinations, utilizing a protocol designed to be completed within five or fewer working days.

M29-A2 **Protection of Laboratory Workers from Occupationally Acquired Infections; Approved Guideline—Second Edition (2001).** Based on U.S. regulations, this document provides guidance on the risk of transmission of hepatitis viruses and human immunodeficiency viruses in any laboratory setting; specific precautions by preventing the laboratory transmission of blood-borne infection from laboratory instruments and materials; and recommendations for the management of blood-borne exposure.

* Proposed- and tentative-level documents are being advanced through the NCCLS consensus process; therefore, readers should refer to the most recent editions.

Active Membership
(as of 1 July 2004)

Sustaining Members

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

Professional Members

American Academy of Family Physicians
American Association for Clinical Chemistry
American Association for Respiratory Care
American Chemical Society
American Medical Technologists
American Society for Clinical Laboratory Science
American Society for Microbiology
American Society of Hematology
American Type Culture Collection, Inc.
Asociacion Mexicana de Bioquimica Clinica A.C.
Assn. of Public Health Laboratories
Assoc. Micro. Clinici Italiani-A.M.C.L.I.
British Society for Antimicrobial Chemotherapy
Canadian Society for Medical Laboratory Science - Société Canadienne de Science de Laboratoire Médical
Canadian Standards Association
Clinical Laboratory Management Association
COLA
College of American Pathologists
College of Medical Laboratory Technologists of Ontario
College of Physicians and Surgeons of Saskatchewan
ESCMID
International Council for Standardization in Haematology
International Federation of Biomedical Laboratory Science
International Federation of Clinical Chemistry
Italian Society of Clinical Biochemistry and Clinical Molecular Biology
Japan Society of Clinical Chemistry
Japanese Committee for Clinical Laboratory Standards
Joint Commission on Accreditation of Healthcare Organizations
National Academy of Clinical Biochemistry
National Association of Testing Authorities - Australia
National Society for Histotechnology, Inc.
New Zealand Association of Phlebotomy
Ontario Medical Association Quality Management Program-Laboratory Service
RCPA Quality Assurance Programs PTY Limited
Sociedad Espanola de Bioquimica Clinica y Patologia Molecular
Sociedade Brasileira de Analises Clinicas
Taiwanese Committee for Clinical Laboratory Standards (TCCLS)
Turkish Society of Microbiology

Government Members

Armed Forces Institute of Pathology
Association of Public Health Laboratories
BC Centre for Disease Control
Caribbean Epidemiology Centre
Centers for Disease Control and Prevention
Centers for Medicare & Medicaid Services
Centers for Medicare & Medicaid Services/CLIA Program
Chinese Committee for Clinical Laboratory Standards
Commonwealth of Pennsylvania Bureau of Laboratories
Department of Veterans Affairs
Deutsches Institut für Normung (DIN)
FDA Center for Devices and Radiological Health
FDA Center for Veterinary Medicine
FDA Division of Anti-Infective Drug Products
Iowa State Hygienic Laboratory
Massachusetts Department of Public Health Laboratories
National Center of Infectious and Parasitic Diseases (Bulgaria)
National Health Laboratory Service (South Africa)
National Institute of Standards and Technology
National Pathology Accreditation Advisory Council (Australia)
New York State Department of Health
Ontario Ministry of Health
Pennsylvania Dept. of Health
Saskatchewan Health-Provincial Laboratory
Scientific Institute of Public Health; Belgium Ministry of Social Affairs,
Public Health and the Environment
Swedish Institute for Infectious Disease Control

Industry Members

AB Biodisk
Abbott Laboratories
Abbott Diabetes Care
Acrometrix Corporation
Advancis Pharmaceutical Corporation
Affymetrix, Inc.
Alifax S.P.A.
Ammirati Regulatory Consulting
Anna Longwell, PC
A/S ROSCO
AstraZeneca Pharmaceuticals
Aventis
Axis-Shield POC AS
Bayer Corporation - Elkhart, IN
Bayer Corporation - Tarrytown, NY
Bayer Corporation - West Haven, CT
BD
BD Consumer Products
BD Diagnostic Systems
BD Thailand Ltd.
BD VACUTAINER Systems
Beckman Coulter, Inc.
Beckman Coulter K.K. (Japan)
Bio-Development SRL
Bio-Inova Life Sciences International
Biomedia Laboratories SDN BHD
bioMérieux, Inc. (MO)
Biometrology Consultants
Bio-Rad Laboratories, Inc.
Bio-Rad Laboratories, Inc. – France
Bio-Rad Laboratories, Inc. – Plano, TX
BioVeris Corporation
Blaine Healthcare Associates, Inc.
Bristol-Myers Squibb Company
Canadian External Quality Assessment Laboratory
Cepheid
Chen & Chen, LLC
Chiron Corporation
ChromaVision Medical Systems, Inc.
Clinical Micro Sensors
The Clinical Microbiology Institute
Cognigen
CONOSCO
Copan Diagnostics Inc.
Cosmetic Ingredient Review
Cubist Pharmaceuticals
Dade Behring Inc. - Cupertino, CA
Dade Behring Inc. - Deerfield, IL
Dade Behring Inc. - Glasgow, DE
Dade Behring Inc. - Marburg, Germany
Dade Behring Inc. - Sacramento, CA
David G. Rhoads Associates, Inc.
Diagnostic Products Corporation
Digene Corporation
Eiken Chemical Company, Ltd.
Elanco Animal Health
Electa Lab s.r.l.
Enterprise Analysis Corporation
EXPERTech Associates, Inc.
F. Hoffman-La Roche AG
Fort Dodge Animal Health
Gen-Probe
GenVault
GlaxoSmithKline
Greiner Bio-One Inc.
Immunicon Corporation
ImmunoSite, Inc.
Instrumentation Laboratory
International Technidyne Corporation
I-STAT Corporation
Johnson and Johnson Pharmaceutical Research and Development, L.L.C.
K.C.J. Enterprises
LAB-Interlink, Inc.
LifeScan, Inc. (a Johnson & Johnson Company)
LUZ, Inc.
Machaon Diagnostics
Medical Device Consultants, Inc.
Merck & Company, Inc.
Minigrip/Zip-Pak
mvi Sciences (MA)
National Pathology Accreditation Advisory Council (Australia)
Nippon Becton Dickinson Co., Ltd.
Nissui Pharmaceutical Co., Ltd.
Norfolk Associates, Inc.
Novartis Pharmaceuticals Corporation
Olympus America, Inc.
Ortho-Clinical Diagnostics, Inc. (Rochester, NY)
Ortho-McNeil Pharmaceutical (Raritan, NJ)
Oxoid Inc.
Paratek Pharmaceuticals
Pfizer Animal Health
Pfizer Inc
Pfizer Italia Srl
Powers Consulting Services
Premier Inc.
Procter & Gamble Pharmaceuticals, Inc.
QSE Consulting
Quintiles, Inc.
Radiometer America, Inc.
Radiometer Medical A/S
Replidyne
Roche Diagnostics GmbH
Roche Diagnostics, Inc.
Roche Laboratories (Div. Hoffmann-La Roche Inc.)
Sarstedt, Inc.
Schering Corporation
Schleicher & Schuell, Inc.
Second Opinion
Seraphim Life Sciences Consulting LLC
SFBC Anapharm
Streck Laboratories, Inc.
SYN X Pharma Inc.
Sysmex Corporation (Japan)
Sysmex Corporation (Long Grove, IL)
TheraDoc
Theravance Inc.
THYMED GmbH
Transasia Engineers
Trek Diagnostic Systems, Inc.
Tyco Kendall Healthcare
Vetoquinol S.A.
Vicuron Pharmaceuticals Inc.
Vysis, Inc.
Wyeth Research
XDX, Inc.
YD Consultant
YD Diagnostics (Seoul, Korea)

Trade Associations

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

Associate Active Members

Academisch Ziekenhuis -VUB (Belgium)
Alfred I. duPont Hospital for Children (DE)
Allina Health System (MN)
American University of Beirut Medical Center (NY)
Anne Arundel Medical Center (MD)
Antwerp University Hospital (Belgium)
Arkansas Department of Health
ARUP at University Hospital (UT)
Associated Regional & University Pathologists (UT)
Atlantic Health System (NJ)
Aurora Consolidated Laboratories (WI)
AZ Sint-Jan (Belgium)
Azienda Ospedale Di Lecco (Italy)
Barnes-Jewish Hospital (MO)
Baxter Regional Medical Center (AR)
Baystate Medical Center (MA)
Bbaguas Duzen Laboratories (Turkey)
BC Biomedical Laboratories (Surrey, BC, Canada)
Bermuda Hospitals Board
Boulder Community Hospital (CO)
Brazosport Memorial Hospital (TX)
Broward General Medical Center (FL)
Cadham Provincial Laboratory (Winnipeg, MB, Canada)
Calgary Laboratory Services (Calgary, AB, Canada)
California Pacific Medical Center
Canterbury Health Laboratories (New Zealand)
Cape Breton Healthcare Complex (Nova Scotia, Canada)
Carilion Consolidated Laboratory (VA)
Carolinas Medical Center (NC)
Cathay General Hospital (Taiwan)
Central Texas Veterans Health Care System
Centro Diagnostico Italiano (Milano, Italy)
Champlain Valley Physicians Hospital (NY)
Chang Gung Memorial Hospital (Taiwan)
Changi General Hospital (Singapore)
Children's Hospital (NE)
Children's Hospital & Clinics (MN)
Children's Hospital Medical Center (Akron, OH)
Children's Medical Center of Dallas (TX)
CHR St. Joseph Warquignies (Belgium)
Christus St. John Hospital (TX)
Clarian Health - Methodist Hospital (IN)
CLSI Laboratories (PA)
Community Hospital of Lancaster (PA)
Community Hospital of the Monterey Peninsula (CA)
CompuNet Clinical Laboratories (OH)
Cook Children's Medical Center (TX)
Cook County Hospital (IL)
Covance Central Laboratory Services (IN)
Creighton University Medical Center (NE)
Danish Veterinary Laboratory (Denmark)
Detroit Health Department (MI)
DFS/CLIA Certification (NC)
Diagnósticos da América S/A (Brazil)
Dr. Everett Chalmers Hospital (New Brunswick, Canada)
Duke University Medical Center (NC)
Dwight David Eisenhower Army Medical Center (GA)
Eastern Health Pathology (Australia)
EMH Regional Medical Center (OH)
Emory University Hospital (GA)
Enzo Clinical Labs (NY)
Evangelical Community Hospital (PA)
Fairview-University Medical Center (MN)
Florida Hospital East Orlando
Focus Technologies (CA)
Focus Technologies (VA)
Foothills Hospital (Calgary, AB, Canada)
Franciscan Shared Laboratory (WI)
Fresno Community Hospital and Medical Center
Gamma Dynacare Medical Laboratories (Ontario, Canada)
Geisinger Medical Center (PA)
Guthrie Clinic Laboratories (PA)
Hagerstown Medical Laboratory (MD)
Harris Methodist Fort Worth (TX)
Hartford Hospital (CT)
Headwaters Health Authority (Alberta, Canada)
Health Network Lab (PA)
Health Partners Laboratories (VA)
Highlands Regional Medical Center (FL)

Hoag Memorial Hospital Presbyterian (CA)
Holy Cross Hospital (MD)
Hôpital du Sacré-Coeur de Montreal (Montreal, Quebec, Canada)
Hôpital Maisonneuve - Rosemont (Montreal, Canada)
Hôpital Saint-Luc (Montreal, Quebec, Canada)
Hospital Consolidated Laboratories (MI)
Hospital for Sick Children (Toronto, ON, Canada)
Hospital Sousa Martins (Portugal)
Hotel Dieu Grace Hospital (Windsor, ON, Canada)
Huddinge University Hospital (Sweden)
Hunter Area Health Service (Australia)
Indiana University
Innova Fairfax Hospital (VA)
Institute of Medical and Veterinary Science (Australia)
International Health Management Associates, Inc. (IL)
Jackson Memorial Hospital (FL)
Jacobi Medical Center (NY)
John C. Lincoln Hospital (AZ)
Johns Hopkins Medical Institutions (MD)
Kadlec Medical Center (WA)
Kaiser Permanente (MD)
Kangnam St. Mary's Hospital (Korea)
Kantonsspital (Switzerland)
Kenora-Rainy River Regional Laboratory Program (Ontario, Canada)
Kimball Medical Center (NJ)
King Faisal Specialist Hospital (Saudi Arabia)
LabCorp (NC)
Laboratoire de Santé Publique du Quebec (Canada)
Laboratorio Dr. Echevarne (Spain)
Laboratório Fleury S/C Ltda. (Brazil)
Laboratorio Manlab (Argentina)
Laboratory Corporation of America (NJ)
LAC and USC Healthcare Network (CA)
Lakeland Regional Medical Center (FL)
Landstuhl Regional Medical Center (APO AE)
LeBonheur Children's Medical Center (TN)
Lewis-Gale Medical Center (VA)
L'Hotel-Dieu de Quebec (Canada)
Libero Instituto Univ. Campus BioMedico (Italy)
Loma Linda Mercantile (CA)
Long Beach Memorial Medical Center (CA)
Louisiana State University Medical Center
Lourdes Hospital (KY)
Maccabi Medical Care and Health Fund (Israel)
Magnolia Regional Health Center (MS)
Maimonides Medical Center (NY)
Marion County Health Department (IN)
Martin Luther King/Drew Medical Center (CA)
Massachusetts General Hospital (Microbiology Laboratory)
MDS Metro Laboratory Services (Burnaby, BC, Canada)
Medical College of Virginia Hospital
Memorial Medical Center (Napoleon Avenue, New Orleans, LA)
Memorial Medical Center (Jefferson Davis Pkwy., New Orleans, LA)
Methodist Hospital (TX)
Michigan Department of Community Health
Mid America Clinical Laboratories, LLC (IN)
Middlesex Hospital (CT)
Monmouth Medical Center (NJ)
Montreal Children's Hospital (Canada)
Montreal General Hospital (Canada)
National Serology Reference Laboratory (Australia)
NB Department of Health & Wellness (New Brunswick, Canada)
The Nebraska Medical Center
New Britain General Hospital (CT)
New England Fertility Institute (CT)
New England Medical Center (MA)
New York University Medical Center
NorDx (ME)
North Carolina State Laboratory of Public Health
North Central Medical Center (TX)
North Shore - Long Island Jewish Health System Laboratories (NY)
North Shore University Hospital (NY)
Northwestern Memorial Hospital (IL)
Ochsner Clinic Foundation (LA)
O.L. Vrouwziekenhuis (Belgium)
Ordre professionnel des technologists médicaux du Québec
Orlando Regional Healthcare System (FL)
Ospedali Riuniti (Italy)
The Ottawa Hospital (Ottawa, ON, Canada)
OU Medical Center (OK)
Our Lady of the Resurrection Medical Center (IL)
Pathology and Cytology Laboratories, Inc. (KY)
Pathology Associates Medical Laboratories (WA)
Peking University Shenzhen Hospital (China)
The Permanente Medical Group (CA)
Piedmont Hospital (GA)
Pocono Medical Center (PA)
Providence Health Care (Vancouver, BC, Canada)
Provincial Laboratory for Public Health (Edmonton, AB, Canada)
Queen Elizabeth Hospital (Prince Edward Island, Canada)
Queensland Health Pathology Services (Australia)
Quest Diagnostics Incorporated (CA)
Quintiles Laboratories, Ltd. (GA)
Regions Hospital
Rex Healthcare (NC)
Rhode Island Department of Health Laboratories
Riverside Medical Center (IL)
Robert Wood Johnson University Hospital (NJ)
Royal Columbian Hospital (New Westminster, BC, Canada)
Sahlgrenska Universitetssjukhuset (Sweden)
St. Alexius Medical Center (ND)
St. Anthony Hospital (CO)
St. Anthony's Hospital (FL)
St. Barnabas Medical Center (NJ)
St-Eustache Hospital (Quebec, Canada)
St. Francis Medical Ctr. (CA)
St. John Hospital and Medical Center (MI)
St. John's Hospital & Health Center (CA)
St. Joseph Mercy Hospital (MI)
St. Joseph's Hospital - Marshfield Clinic (WI)
St. Joseph's Hospital & Medical Center (AZ)
St. Jude Children's Research Hospital (TN)
St. Mary of the Plains Hospital (TX)
St. Michael's Hospital (Toronto, ON, Canada)
Ste. Justine Hospital (Montreal, PQ, Canada)
Salem Clinic (OR)
San Francisco General Hospital (CA)
Santa Clara Valley Medical Center (CA)
Seoul Nat'l University Hospital (Korea)
Shands at the University of Florida
South Bend Medical Foundation (IN)
South Western Area Pathology Service (Australia)
Southern Maine Medical Center
Southwest Texas Methodist Hospital
Spartanburg Regional Medical Center (SC)
Specialty Laboratories, Inc. (CA)
State of Connecticut Dept. of Public Health
State of Washington Department of Health
Stony Brook University Hospital (NY)
Stormont-Vail Regional Medical Center (KS)
Sun Health-Boswell Hospital (AZ)
Sunnybrook Health Science Center (ON, Canada)
Swedish Medical Center - Providence Campus (WA)
Temple University Hospital (PA)
Tenet Odessa Regional Hospital (TX)
The Toledo Hospital (OH)
Touro Infirmary (LA)
Tripler Army Medical Center (HI)
Truman Medical Center (MO)
UCLA Medical Center (CA)
UCSF Medical Center (CA)
UNC Hospitals (NC)
Unidad de Patologia Clinica (Mexico)
Union Clinical Laboratory (Taiwan)
Universita Campus Bio-Medico (Italy)
University College Hospital (Galway, Ireland)
University Hospitals of Cleveland (OH)
University of Alabama-Birmingham Hospital
University of Chicago Hospitals (IL)
University of Colorado Hospital
University of Debrecen Medical Health and Science Center (Hungary)
University of Illinois Medical Center
University of the Ryukyus (Japan)
The University of Texas Medical Branch
The University of the West Indies
University of Virginia Medical Center
University of Washington
UZ-KUL Medical Center (Belgium)
VA (Hampton) Medical Center (VA)
VA (Hines) Medical Center (IL)
VA (Tuskegee) Medical Center (AL)
Valley Children's Hospital (CA)
Vejle Hospital (Denmark)
Virginia Beach General Hospital (VA)
Virginia Department of Health
Virginia Regional Medical Center (MN)
ViroMed Laboratories (MN)
Washington Adventist Hospital (MD)
Washoe Medical Center Laboratory (NV)
Waterford Regional Hospital (Ireland)
Wellstar Health Systems (GA)
West Jefferson Medical Center (LA)
Wilford Hall Medical Center (TX)
William Beaumont Army Medical Center (TX)
William Beaumont Hospital (MI)
William Osler Health Centre (Brampton, ON, Canada)
Winn Army Community Hospital (GA)
Winnipeg Regional Health Authority (Winnipeg, Canada)
Wishard Memorial Hospital (IN)
Yonsei University College of Medicine (Korea)
York Hospital (PA)

NCCLS ▼ 940 West Valley Road ▼ Suite 1400 ▼ Wayne, PA 19087 ▼ USA ▼ PHONE 610.688.0100
FAX 610.688.0700 ▼ E-MAIL: exoffice@nccls.org ▼ WEBSITE: www.nccls.org ▼ ISBN 1-56238-542-9