EP6-A
Vol. 23 No. 16
Replaces EP6-P2
Vol. 21 No. 20

Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline

This document provides guidance for characterizing the linearity of a method during a method evaluation; for checking linearity as part of routine quality assurance; and for determining and stating a manufacturer's claim for linear range.

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

NCCLS...

Serving the World's Medical Science Community Through Voluntary Consensus

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 (i.e., "tentative") 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.

Tentative A tentative standard or guideline is made available for review and comment only when a recommended method has a well-defined need for a field evaluation or when a recommended protocol requires that specific data be collected. It should be reviewed to ensure its utility.

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.

EP6-A
ISBN 1-56238-498-8
Volume 23 Number 16 ISSN 0273-3099

Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline

Daniel W. Tholen, M.S., Chairholder
Martin Kroll, M.D., FACB
J. Rex Astles, Ph.D., FACB
Albert L. Caffo, Ph.D.
Thomas M. Happe, MT(ASCP)
Jan Krouwer, Ph.D.
Fred Lasky, Ph.D.

Abstract

NCCLS document EP6-A— *Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline* is intended to provide both manufacturers and users of quantitative analytical methods with an economical and user-friendly method of establishing or verifying the linear range. This guideline also can be used to demonstrate the extent to which a quantitative analytical method meets clinical requirements or manufacturer's linear range claims.

NCCLS. *Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline*. NCCLS document EP6-A (ISBN 1-56238-498-8). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2003.

NCCLS

Suggested Citation

(NCCLS. *Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline.* NCCLS document EP6-A [ISBN 1-56238-498-8]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2003.)

Proposed Guideline
October 1986

Proposed Guideline—Second Edition
December 2001

Approved Guideline
April 2003

ISBN 1-56238-498-8
ISSN 0273-3099

Committee Membership

Area Committee on Evaluation Protocols

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

Daniel W. Tholen, M.S.
Vice-Chairholder
Dan Tholen Statistical Services
Traverse City, Michigan

Working Group on Linearity

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

Martin H. Kroll, M.D., FACB
VA North Texas Healthcare System
Dallas, Texas

J. Rex Astles, Ph.D., FACB
Centers for Disease Control and Prevention
Atlanta, Georgia

Albert L. Caffo, Ph.D.
West Chester University
West Chester, Pennsylvania

Thomas M. Happe, M.T. (ASCP)
Maine Standards Company
Portland, Maine

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

Fred D. Lasky, Ph.D.
Ortho Clinical Diagnostics
Rochester, New York

Staff

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

Patrice E. Polgar
Project Manager, Editor
NCCLS
Wayne, Pennsylvania

Donna M. Wilhelm
Assistant Editor
NCCLS
Wayne, Pennsylvania

Contents

Contents (Continued)

Foreword

A quantitative analytical method is linear when there exists a mathematically verified straight-line relationship between the observed values and the true concentrations or activities of the analyte. The property of being linear is important for analytical and clinical laboratory methods. The linear relationship is valuable because it represents the simplest mathematical relationship and allows for simple and easy interpolation of results.

It has been argued that linearity is not necessary for such methods as competitive immunoassays, because the relationship between the response of the system and analyte concentration is inherently nonlinear. However, the mathematical relationship between the response and analyte concentrations should be sufficiently well defined to allow the selection of a suitable transformation of the dose-response curve to linear form. Furthermore, for an analytical method, one cannot interpolate between points unless one knows the results are linear.

The reason for ensuring linearity of analytical methods is important on clinical grounds. Clinicians know that the relationship between an analyte and the pathophysiologic process is usually nonlinear, but they expect that the results reported to them by the laboratory include a linear relationship between the result and the true concentration, count, or activity recovered. For example, if the true amount of an analyte in a sample were to double, clinicians expect that would be reflected in a doubling of the measured value.

This Revision

The approach outlined within the approved version of EP6 represents a significant change in statistical approach from the previous version.

The first edition (EP6-P) used the statistical "Lack of Fit" test (LoF) as the sole basis for determining linearity. With this protocol, five equally spaced concentrations were analyzed with four replicates at each level. A regression line was fit to the points, and two variance estimates were produced – the pooled variance between replicates and the variance of the five means around the regression line. The ratio of these variances was the basis for the LoF test. One problem with this approach is that very precise systems have small variance between replicates, and in the ratio LoF test, very small deviations of the means were seen as statistically significant – or nonlinear. It was also possible – though less frequent – that extremely large variance between replicates could lead to a failure to detect clinically important nonlinearity.

This polynomial method (EP6-A) is conceptually very similar to the original LoF test. In both procedures, there are two alternative statistical models (linear and nonlinear); and they are assessed relative to which is most likely to be true. However, with the polynomial approach, there is a specific parametric model for the nonlinear alternative, to better identify specific nonlinear conditions, and pure repeatability precision is included in a positive way. Better precision leads to better decisions and poor precision is screened out before linearity decisions are made. This approved statistical approach:

- estimates the magnitude of nonlinearity at every level;
- controls for unacceptable repeatability in the assessment of nonlinearity;
- provides a testable statistical model;
- can be programmed easily with widely available software.

Another key concept from the previous version has been retained, i.e., the necessity to plot the data as a first step in the evaluation of linearity. The visual examination will help the user decide the extent to which the statistical assessment can be useful for determining whether there is significant nonlinearity, and the magnitude of the nonlinearity relative to predetermined goals.

A Note on Terminology

NCCLS, as a global leader in standardization, is committed to achieving global harmonization wherever possible. Harmonization is a process of recognizing, understanding, and explaining differences 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 is an area of immediate attention. Implementation of this policy is an evolutionary and educational process that begins with new projects and revisions of existing documents.

In keeping with NCCLS's commitment to align terminology with that of ISO, the following terms are used in EP6: *Trueness* is used in this document when referring to the closeness of the agreement between the average value from a large series of measurements and to a true value of a measurand; *Repeatability* has replaced the term *Within-run precision* where appropriate, when describing the closeness of agreement between results of successive measurements of the same measurand carried out under the same conditions of measurement; *Measurement procedure* has replaced the term *Analytical method* for a set of operations, used in the performance of particular measurements according to a given method; *Measuring range* has replaced *Reportable range* when referring to a set of values of measurands for which the error of a measuring instrument is intended to lie within specified limits; *Measurement error/Error of measurement* is used instead of *Total error* to describe the result of a measurement minus a true value of the measurand.

At this time, the working group has chosen not to replace A*nalyte* with *Measurand,(i.e.,* particular quantity subject to measurement) due to user nonfamiliarity and for the sake of the practicability of the guideline.

Users of EP6-A should understand, however, that the fundamental meanings of the terms are very similar, and to facilitate understanding, the terms are defined along with explanatory notes in the guideline's Definitions section.

All terms and definitions will be reviewed again for consistency with international use, and revised appropriately during the next scheduled revision of this document.

Standard Precautions

Because it is often impossible to know what might be infectious, all human blood specimens are to be treated as infectious and handled according to “standard precautions.” Standard precautions are new guidelines that combine the major features of “universal precautions and body substance isolation” practices. Standard precautions cover the transmission of any pathogen and thus are more comprehensive than universal precautions which are intended to apply only to transmission of blood-borne pathogens. Standard precaution 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;Vol 17;1:53-80), (MMWR 1987;36[suppl 2S]2S-18S), and (MMWR 1988;37:377-382, 387-388). For specific precautions for preventing the laboratory transmission of blood-borne infection from laboratory instruments and materials and for recommendations for the management of blood-borne exposure, refer to the most current edition of NCCLS document M29—*Protection of Laboratory Workers from Occupationally Acquired Infections.*

Key Words

Allowable difference, allowable error, linearity, matrix effects, measurement error, total error, uncertainty

Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline

1 Scope

This document presents a method to establish, verify and/or demonstrate the linear range of a quantitative measurement procedure. These methods do not identify the causes of significant nonlinearity. The method employs increasing numbers of samples for more definitive examinations of linearity. Therefore, if a failed demonstration is evaluated and it is determined that the experiment needs to be repeated, it can be done with more replicates or with fewer levels to cover a smaller range.

- This protocol is to assess linearity, isolated as much as possible from conditions of precision and trueness. It is understood that poor precision will hinder an effective assessment of linearity, so a check for poor repeatability is included.

- These experiments should use samples with a matrix appropriate to the specimens being analyzed (serum, plasma, urine, etc.)

- This protocol requires laboratories to set goals for nonlinear error. It provides basic concepts for setting such goals, but does not recommend any specific protocol.

2 Introduction

2.1 Purpose

The purpose of this guideline is to describe a statistical process for determining the linearity of a quantitative measurement procedure. This primary objective is to determine the concentration(s) where a method is not linear and the extent of the nonlinearity at that level. This guideline emphasizes the necessity that each user establishes his or her requirements for linearity, or the allowable error due to nonlinearity. It also places less importance on global tests for linearity across the tested range (such as the LoF test). Global tests merely indicate that statistically significant nonlinearity exists; they do not show where that nonlinearity is, nor do they show the magnitude of the error. Linearity tests can be helpful to assess bias, which is a component of measurement error, but nonlinearity is not the only component of bias.

Users should have an understanding of their needs for measurement error, bias, random error (or imprecision), and nonspecificity (or interferences). From these they can derive a goal for linearity. In this context, NCCLS document EP6-A is a part of a series of documents that guide users through the process of method evaluation. Also see the most recent version of NCCLS document EP21—*Total Analytical Error for Clinical Laboratory Methods*.

This document is meant to cover a broad range of situations, such as establishing the linear range of a method, which requires testing across a wide range of concentrations, then progressively narrowing into a range of acceptable linearity. It is also intended to cover situations where the linear range has been determined elsewhere (e.g., by the manufacturer), but the user wishes to verify that range in their laboratory. The same procedure is used for all scenarios, but with different numbers of concentration levels and different numbers of replicates.

2.2 Alternative Approaches

There are many experimental approaches used to evaluate claims of a linear range. Kroll and Emancipator[1,2,3] proposed a method based on comparison of a linear regression equation to a series of polynomial equations, using the residuals (distances from each data point to the calculated regression line) to define linearity. The advantages of this method are that it is robust and based on rigorous scientific principles. Demonstration that the linear (first-order) fit is the best provides evidence that the method is linear.

The process is not simple, because users must be able to test a variety of polynomial equations against the experimental data (although spreadsheet programs can be used). These procedures were further refined by Kroll, Praestgard, Michaliszyn, and Styer[4] and implemented in the College of American Pathologists interlaboratory comparison programs. Krouwer and Schlain[5] proposed a general method using the "last point off the line" (LPO) method, which requires iterative calculations of a multiple least-squares equation. Tholen[6] proposed a graphic evaluation along with a regression lack of fit compared to a goal percentage; and alternatively a simple analysis of the slopes of consecutive line segments.

2.3 Definitions

Allowable difference/Allowable error - The magnitude of analytical deviation, from all sources, that a user can tolerate in a testing system and still meet the medical requirements of the test; **NOTE*:*** The allowable difference (error) boundaries (for a single observation) are represented by the target value of the specimen plus or minus the allowable error amount.

Analyte - Component indicated in the name of a measurable quantity (***ISO FDIS 17511).***[7]

Analytical result - For this document, the final result of a measurement on a test specimen; **NOTE:** The result is usually in concentration or activity units; it is assumed that the measurement procedures to be evaluated by the procedures in this guideline are quantitative methods that yield a numerical output.

Bias - The difference between the expectation of the test results and an accepted reference value;[8] **NOTE:** In general, the deviation/difference is based on replicate measurement using an accepted (definitive, reference, or designated comparison) method and the method being tested, and expressed in the units of the measurement or as a percentage.

Input variable - The given value that is used as a reference (independent variable) and against which the output variable is compared; **NOTE:** The input variable is represented by X with individual values of X noted by x_i (i represents an individual observation); its value is plotted along the X-axis (abscissa).

Intercept//Y intercept - 1) The point where a function intersects an axis; **2)** The value of a variable, when the value for the other variables is zero; **NOTE:** The *y*-intercept is the value of the *y*-variable when the *x*-variable has a value of zero.

Least squares regression - The method of statistically placing the location of the estimated line or curve among the data so that the sum of the squares of the distances of each data point from the line in the perpendicular direction from the X-axis (i.e., parallel to the Y-axis) is minimized; **NOTE:** It allows the direct algebraic computation of the coefficients and an estimate of their uncertainty.

Linear equation - An equation that represents a straight line; **NOTE:** A linear equation is typically represented mathematically by: Y = a + bX; X and Y are the input and output variables, b is the slope, and a is the y-intercept.

Linear range - For this document, the range over which the testing systems results are acceptably linear; that is, where nonlinear error is less than the error criterion.

Linearity - The ability (within a given range) to provide results that are directly proportional to the concentration {amount} of the analyte in the test sample*;* **NOTES:** a) Linearity typically refers to overall system response (i.e., the final analytical answer rather than the raw instrument output**;** b) The linearity of a system is measured by testing levels of an analyte which are known by formulation or known *relative to each other* (not necessarily known absolutely); when the system results are plotted against these values, the degree to which the plotted curve conforms to a straight line is a measure of system linearity.

Linear regression - A statistical calculation that results in parameters that describe the assumed linear relationship between values of an independent and a dependent variable wherein the independent variable is known exactly; **NOTE:** a) The calculation is based on the mathematical definition of a line *(*Y = a +bX); and the mathematical minimization of the vertical distance between each data point and the regression line.

Measurement error//Error of measurement - The result of a measurement minus a true value (or accepted reference value) of the measurand;[9] **NOTE:** a) Formerly, the term **Total error,** was used in this document.

Measurement procedure - A set of operations, described specifically, used in the performance of particular measurements according to a given method;[9] **NOTE:** Formerly, the term **Analytical method** was used in this document.

Measuring range - A set of values of measurands for which the error of a measuring instrument is intended to lie within specified limits;[9] **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, the term **Reportable range** was used in this document.

Output variable - The dependent variable, the value of which is compared to the input (independent) variable; **NOTE:** The output variable is represented by Y, with individual values of Y noted by y_i (i represents an individual observation), and its value is plotted along the Y axis (ordinate); the mean of all the output variables (y_i) in the data pool is $\bar{y}$.

Polynomial regression – A least-squared regression using polynomials of various orders:

$Y = a + b_1x$	(first-order polynomial or linear fit)
$Y = a + b_1X + b_2X^2$	(second-order polynomial), and
$Y = a + b_1X + b_2X^2 + b_3X^3$	(third-order polynomial)

Repeatability{/Repeatability of a measuring system/instrument} - The ability of a measuring {system/}instrument to provide closely similar indications for repeated applications of the same measurand under the same conditions of measurement;[9] **NOTES:** a) These conditions include reduction to a minimum of the variations due to the observer; same measurement procedure; same measuring equipment used under the same conditions, same location; and repetition over a short period of time; b) Repeatability may be expressed quantitatively in terms of the dispersion characteristics of the indications.

Standard deviation of *y* about a regression//Standard error of regression ($SD_{y,x}$) - A measure of the dispersion of the observed values of the dependent variable Y about the estimated average value of this variable $\hat{Y}$ over the various values of the predictor variable X.

Testing system - Includes the entirety of the testing process, including instrument, sample, personnel, reagents, supplies, and procedures; **NOTE:** It also includes such attributes (method characteristics) as sample type, testing temperature, humidity, etc.

Trueness - The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value;[8] **NOTE**: Trueness is usually expressed numerically by the statistical measure bias that is inversely related to trueness. See also **Bias**.

2.4 Overview of the Experimental Approach

This guideline uses the polynomial method originally proposed by Kroll et al.,[3,4,5] because it is statistically rigorous and can be easily implemented by spreadsheet programs. These procedures are based on conventional statistical methods as described in the above documents, in basic statistical texts[10,11,12] and in conventional statistics software.

The polynomial evaluation of linearity first assumes that the data set is nonlinear. This approach is parametric, i.e., it assumes that the data points can be described by a mathematical equation describing a line or curve, and that the random error is normally distributed. Whether the best-fit curve is linear or not does not affect the ability to interpolate between the experimental points.

The basic data collection requires multiple measurements from five to nine samples with varying concentrations, which are known relative to one another by dilution ratios or by formulation. There is no requirement for the sample concentrations to be equidistant; and there is no requirement for the assumed values to be obtained by dilution.

In essence, the polynomial method evaluates for nonlinearity, since polynomials are designed to fit curved responses. The method consists of two parts. The first part examines whether a nonlinear polynomial fits the data better than a linear one. The second part assesses whether the difference between the best-fitting nonlinear and linear polynomial is less than the amount of allowable bias for the method.

3 Time Requirements

3.1 Device Familiarization Period

The laboratory personnel performing the analytical procedure must be familiar with the system operation, ensure it is in control and correctly calibrated, and that samples are prepared properly. Appropriate manufacturer's training programs, when offered, are a useful part of the familiarization period. The instrument should be set up and operated in the laboratory long enough to ensure that the laboratory personnel understand the procedure necessary for proper operation.

3.2 Duration of the Experiment

The entire data gathering protocol should take place in as short a time interval as possible after the device familiarization period has been completed. Ideally, all results for a single analyte will be obtained on the same day.

4 Evaluating Experimental Conditions

The experimental conditions required to establish the linear range may be different than protocols to demonstrate that the range is appropriate for a particular testing system. Establishing a range may take more replicates and perhaps a larger number of concentrations. The demonstration of the linear range

requires a series of known concentrations or known relationships established by dilution. In all testing situations, the testing sequence should be random.

4.1 Specimen Requirement

The linear range experiment requires enough of each specimen to prepare dilutions and to carry out the testing for each of five (or more) concentrations. The volume can vary by setting:

- Laboratories needing to demonstrate that a system's operation is within the claimed performance should use 5 to 7 concentrations chosen throughout the stated linear range. Laboratories should test 2 replicates at each level.

- Developers of new methods who are seeking to establish linear ranges (including manufacturers and laboratories that modify existing methods) should use 7 to 11 concentrations across the anticipated measuring range. Such developers may wish to use more points and a range that is 20 to 30% wider than the anticipated measuring range, and plan to drop off points to discover the widest possible linear range. Experiments should use 2 to 4 replicates at each level, depending on the expected imprecision.

- Five points are the minimum number to reliably describe the linear range and use the polynomial method. More points will provide a more exact description of linearity and will allow a wider linear range, should points need to be deleted.

It is recommended that equally spaced intermediate concentration levels be prepared accurately by proportionately mixing the high and low pools (see Sections 4.6 and 4.7), although the proposed method does not require equally spaced concentrations. It is acceptable to use samples especially prepared for specific concentrations, as long as the concentrations are known relative to each other. For the purpose of this evaluation, accurately pipetted mixtures of high and low pools are less likely to produce errors than individually weighed specimens or reconstituted solutions prepared in the laboratory. Caution should be exercised in pipetting small volumes. Positive displacement pipettes are useful in preparing accurate dilutions with small volumes. Dilution schemes for preparing samples with 5 to 11 equally spaced concentrations are provided in Appendix A.

Commercial products intended for *in vitro* diagnostics may be used, providing manufacturer instructions are followed. These products can be used to quantitatively verify calibration, validate measuring ranges, and determine linearity in automated, semiautomated, and manual chemistry systems.

4.2 Matrix Effects

The specimens used to determine the linear range of the analytical output should be similar to the specimens that will be used in clinical testing. All specimens should be free of known interferences that the method instructions identify (such as that which may be seen with icterus, hemolysis, or lipemia). If this condition cannot be met, the final report should state any specimen treatment or matrix type used in the evaluation. **Caution:** Sample treatments can have different effects on adulterants than on human specimens and can create interferences.

4.3 Hierarchy of Acceptable Matrices

If multiple matrix specimen types (urine, serum, spinal fluid, whole blood, etc.) are to be assayed, a study should be carried out for each specimen type. For additional information, refer to the most current version of NCCLS document EP14— *Evaluation of Matrix Effects.*

The following is a list of possible specimen types in order of their desirability. Care should be taken by using the appropriate matrix for the particular method.

4.3.1 Patient-Sample Pool

The ideal sample matrix is a patient's specimen pool with an analyte concentration near the expected upper reportable limit that is diluted with another patient's sample pool having an analyte concentration at the expected or tested lower limit. Note that the final concentration of the pools that are actually analyzed represents the tested range. Therefore, the final high and/or low pools may require adjustment to bring them into the desired range.

4.3.2 Patient-Sample Pool Diluted With a Recommended Diluent

The fluid chosen to dilute a high concentration patient-sample pool can have an effect on the test results. Only the diluent recommended by the manufacturer or proven acceptable by the laboratory should be used for dilution. For those analytical systems that allow specimen dilution for patient specimens having analyte concentrations above the measuring range, using the approved diluent to demonstrate linearity across the measuring range has the added benefit of documenting that the diluent is appropriate. Dilution studies should be included in the manufacturer's establishment of performance, as appropriate.

4.3.3 Patient-Sample Pool Supplemented With Analyte

The supplemental material containing the analyte need not be of high purity if interfering substances are absent. If interfering substances are present, they should be noted (source, purity, and anticipated effect) in the report. If a concentrated solution of analyte is used to supplement the pool, dilute the patient's sample pool as little as possible (less than 10%, as a guide), and document the solvent.

4.3.4 Pool Diluted With Treated Materials of Low Concentration or Treated Pool Material

If possible, use a low concentration patient-specimen pool. Alternatively, certain treatments can be used to reduce the analyte concentration, (e.g., dialysis, heat treatment, and chromatography). Be aware that these treatments may change the analyte and/or the matrix physically and/or chemically. It is important to retain a constant matrix rather than achieve low levels simply by dilution.

4.3.5 Commercial Control/Calibrator/Linearity Material

If these materials are used, assay them as you would assay patient specimens. If appropriate admixtures are not provided, use the high and low samples to create intermediate concentrations. Be sure that the analyte is in a physiologically normal form, (e.g., protein-bound or a metabolite, if appropriate).

4.3.6 Pool Diluted With Saline or Other Diluents Other Than the Recommended Diluent.

When using such samples, be cautious of matrix differences that can affect the results. These dilutions should be kept to a minimum and documented.

4.3.7 Underdiluted / Overdiluted Commercial Control Material

When using such samples, be cautious of matrix differences that can affect the results: these effects can vary by concentration. Dilutions should be kept to a minimum and documented. Care should be taken to achieve complete solution.

4.3.8 Aqueous Solutions

Human matrix effects are not tested for their effects on the response of the method when aqueous solutions are used. Matrix effects may influence the response of the method and the interpretation of results. Although materials of high purity are preferable since they minimize the possibility of interferences that could affect interpretation of the results, materials of lesser purity could be acceptable. A large number of routine clinical chemistry methods are calibrated with aqueous materials, and the weigh-in concentration can be accepted as the target value.

4.3.9 Solution in Other Solvents

Matrix effects are even more likely when organic solvents are used.

4.4 Selection of Materials Used to Supplement Samples

Caution should be exercised when selecting these materials. Effects can arise from the method of preparation for some constituents (such as using enzymes from other species).

Table 1 lists suggested material sources for several analytes. The instrument or reagent manufacturer may also be of help in determining suitable materials.

Table 1. Materials for Spiking Samples

Analyte	Material
Alanine aminotransferase (ALT)	Purified enzyme
Albumin	Powdered albumin, human fraction V
Alcohol	Stock ethanol
Alkaline phosphatase	Purified enzyme, if elevated patient sample is unavailable
Aspartame aminotransferase (AST)	Purified enzyme
Amylase	Saliva or pancreatic extract
Bilirubin	Purified bilirubin, or underdiluted elevated control, or standard if elevated patient sample is unavailable
Cholesterol	Purified cholesterol, or underdiluted control
Carbon dioxide (CO_2)	Sodium carbonate, Sodium bicarbonate
Creatine kinase (CK)	Purified enzyme
Creatinine	Stock standard or elevated patient sample
g-Glutamyl-transferase	Purified enzyme
Hemoglobin	Human washed and lysed red cells
Hematocrit	Microhematocrit-spun human red cells
Lactate dehydrogenase (LDH)	Purified enzyme
Lipase	Pancreatic extract
Magnesium	Magnesium chloride
Phosphorus	Potassium or sodium phosphate monobasic
Total protein	Powdered albumin, human (preferred) or bovine fraction V. (A representative total protein material is very difficult to prepare because of its many constituents. If albumin is substituted for total protein or a nonhuman species is used, this fact should be noted in the experimental details and in the results).
White blood cells	Human whole blood buffy coat

NOTE: The source of materials, especially proteins and enzymes, should be noted to ensure that differences in physical and/or chemical reactivity are documented.

Solubility must be considered when using some of the recommended raw materials. Pure cholesterol, for example, is not readily soluble in serum or aqueous matrices. Organic solvents are required for dissolution. Raw-material enzyme sources must be carefully chosen to ensure that enzyme activity is equivalent to that of human source materials, and that additives do not affect the results. While human source enzymes might be the most appropriate choice, they are often expensive, if available at all. Albumin recovery using sources other than human albumin is often compromised, depending on the dye-binding method used and its source of calibration material. Certified reference materials, such as chemicals from NIST and ACS are appropriate for many of the common analytes such as glucose, urea, sodium, chloride, etc.

4.5 Analytical Sequence

The analytical sequence should be random.[13] If significant carry-over or drift is present, the experiment may be compromised.

4.6 Analyte Range

The concentration levels chosen for measurement should encompass or be equal to the minimum and the maximum values that are specified in the performance claims. If the linear range or measuring range claims do not hold for the concentrations chosen, a new analysis should be done with more appropriate concentrations; or if five or more analyte concentrations are analyzed, the data from the terminal pool (if appropriate) can be eliminated and the range appropriately shortened.

When establishing the linear range, use 7 to 11 levels over a range that is 20 to 30% wider than the anticipated measuring range, and then plan to eliminate nonlinear points to establish the widest possible range of acceptable linear response. Using more levels can help justify the use of data interpolation to define performance over the linear range.

4.7 Sample Preparation and Value Assignment

If the concentration of the high and/or low pools is not known, each pool can be coded. Coding is the process of assigning each pool a number that conveys its relative concentration. For equally spaced concentrations, the coding allows assignment of whole numbers (i.e., 1, 2, 3, 4, and 5) to each consecutive pool. In other words, the pool concentrations do not need to be known ahead of time. To verify the limits, the mean of the results for the high and low pools can be used. If intermediate analyte concentrations are not equally spaced, the concentration interval relationship should be known and these intervals used as the values of X in the analysis. These analyte concentrations can also be made by preparing a midlevel sample from the low and high pools and preparing intermediate concentrations from the low and mid pools and the mid and high pools.

An example of a process for five equally spaced samples is provided below. This example uses a dilution-based scheme. In some laboratories, a gravimetric scheme may be preferred. Regardless of the procedure for preparing solutions, the preparer should use due caution. Appendix A gives similar dilution schemes for 5 to 11 equally spaced samples:

(1) Two pools (one high- and one low-concentration), each in an acceptable matrix (see Sections 4.2 to 4.4), are obtained in volumes sufficient for the entire experiment. The required volume of each pool will depend on the assay volume requirement. Care should be taken to use the most accurate and precise pipetting technique available. The following pool designations are based on equally spaced concentrations.

(2) The low-concentration pool (ideally near and within the lower limit) is coded Pool No. 1; the high-concentration pool (the highest concentration tested) is coded Pool No. 5.

(3) The intermediate concentration pools are made by dilutions that will be related to each other and the high and low pools by constant intervals. A convenient method for making the intermediate pools is as follows:

Pool 2 is a mixture of three parts of the low pool (No. 1) and one part of the high pool (No. 5).

Pool 3 is a mixture of two parts of the low pool (No. 1) and two parts of the high pool (No. 5).

Pool 4 is a mixture of one part of the low pool (No. 1) and three parts of the high pool (No. 5).

The concentration of each pool is defined by the following formula, in which concentration of Pool No. 1 is C1 and the volume part of Pool No. 1 is V1; likewise, the concentration of Pool No. 5 is C5 and the volume part of Pool No. 5 is V5:

$$\text{Concentration} = (C1 \cdot V1 + C5 \cdot V5)/(V1+V5)$$

All concentration and volume units must be the same for each pool. Care must be taken to mix each pool thoroughly and to protect the pools from evaporation or other deterioration.

For example, for a testing system that requires 0.05 mL per test, and n=10 per level is tested, at least 0.5 mL plus dead volume will be needed. The high concentration pool is 120 units and the low concentration pool is 40 units. **NOTE:** The concentration does not need to be known before testing.

(1) Specimens for Pools 1 and 5, 0.6-mL aliquots are placed separately in appropriately marked sample tubes for assay.

(2) Pool 2 is made by adding 0.600 mL of Pool 1 to 0.200 mL of Pool 5. The resultant pool is labeled Pool 2.

(3) Pool 3 is made by combining 0.400 mL of Pool 1 and 0.400 mL of Pool 5.

(4) Finally, Pool 4 is made by combining 0.200 mL of Pool 1 and 0.600 mL of Pool 5. For example, the expected concentration of Pool 4 is calculated as:

$$\text{Pool 4} = (40 \cdot 0.200 + 120 \cdot 0.600)/(0.200 + 0.600) = 100 \text{ units.}$$

The expected concentrations of the pools are 40, 60, 80, 100, and 120 units.

The plotting of the analytical results can use known concentrations, the coded pool numbers (i.e., 1,2,3,4, and 5), or the calculated concentrations.

If other concentration relationships are required, they can be used, but the concentration interval relationship should be known and these intervals appropriately used in the data plots. If it were required, for example, to have a concentration of 110 units in the above example, it could be made by combining 0.100 mL of Pool 1 and 0.700 mL of Pool 5, and its coded pool number would be 4.5. Similar calculations could be used to interpolate from any set of concentrations.

The examples in Appendix C demonstrate the use of coded concentrations as outlined above.

4.8 Data Collection

Data can be conveniently listed on data worksheets in the laboratory or in computer spreadsheets.

The following are the numbers of samples suggested for different purposes.

- To establish the linear range: 9 to 11 levels, and 2 to 4 replicates at each level are recommended.
- To validate claims for an "in-house" or modified method: 7 to 9 points, and 2 or 3 replicates at each level are recommended.
- To confirm that the linear range is valid in a laboratory: 5 to 7 levels, and 2 replicates at each level are recommended.

The samples should be assayed randomly during a single run or closely grouped analytical runs. Any data elimination will be handled as described in Sections 5.1 and 5.2.

The number of replicates should be sufficient to produce a reliable estimate of the concentration at each level. For some analytes, at some concentrations, that may require 3-5 replicates. Users should use their best judgment about the number of replicates needed. The procedures in this protocol are not adversely affected by the use of different numbers of replicates at different levels.

4.9 Clinically Important Evaluation Concentrations

The evaluation of the linear range should include the following important concentrations:

- minimum analytical concentration or lower limit of the linear range;
- various medical decision limits; and
- maximum analytical concentration or the upper limit of the linear range.

5 Demonstration of Linear Range—Data Evaluation

5.1 Preliminary Data Examination

The experimental results should be evaluated for general acceptability and usefulness. If the data appear to be unusable, refer to Section 6.3 for a discussion on setting error goals. A detailed evaluation of the results should use the following stepwise pattern:

(1) Examine the recorded data for obvious excessive differences (errors). If an analytical or technical problem can be identified and corrected, the linear range can be assessed by repeating the entire set of specimens.

(2) If analytical or technical excessive differences (errors) are not in evidence, visually examine the plotted data for potential outliers at each analyte concentration. This includes making a plot of the results with the response variable (**Y**) on the vertical axis and the sample concentration (or relative concentration, **X**) on the horizontal axis. To do this, plot the individual results for each result (y) at each known or relative value of the sample (x); optionally, plot the mean value for each set of replicates at each level. If means are plotted, connect the points either manually or with the computer program. Observe the plot for gross deviations from linearity, misplaced points, obvious transcription errors, or evidence of instrument failures.

(3) If needed, arrange the data for each pool in chronological order to evaluate for drift or trends. If any excessive difference (error) is found, the entire run's data should be replaced after the cause is corrected. Be careful not to search for "good" data by repeating the test result multiple times without correction of the cause. Data showing a problem may be true indications of performance.

(4) Look at the differences between responses at each level. In a linear pattern, the segment slopes will be approximately equal. Any increasing or decreasing trend is an indication of nonlinearity.

(5) If a response value (y_i) seems too far removed from the other Y values for a given concentration, visually evaluate it as an outlier (Section 5.2). Any response meeting the outlier criterion should be eliminated from the data set.

(6) Two or more unexplained outliers casts doubt on the testing system's performance. Conduct troubleshooting procedures, which may include contacting the manufacturer for assistance if these outliers are unexplained.

(7) A visual examination of the XY plot is important for guiding the subsequent assessment of linearity. This will show whether there is obvious nonlinearity, or if the range of testing should be narrowed or expanded. It will also give insight into the most appropriate statistical procedures for the subsequent statistical analysis.

5.2 Outlier Test

In this document, "outliers" are single results that are visually and/or statistically different than other results, and apply only to single replicate values. They are not multiple replicates at the same level or an average value at one level; that type of deviant result would be an indication of nonlinearity or other systematic error. Outliers are test results that come from mistakes (clerical errors, system glitches, etc.), or from assumed mistakes, whether proven or reasonably assumed. For further discussion, see ASTM E 178-94.[14] Outliers are results that do not fit the pattern represented by the rest of the data. There are a number of statistical outlier tests; however the most sensitive test for outliers is the visual evaluation of points from a graph of results versus expected values. A single outlier in a dataset can be removed and does not need to be replaced. If more than one point appears to meet the subjective evaluation criteria, the testing system is probably imprecise, and the values should be taken as typical operation. In this case the cause should be identified and, if possible, the cause should be eliminated.

5.3 Determination of the Linear Range

5.3.1 Summary of Methods—The Polynomial Evaluation of Linearity

The polynomial evaluation of linearity first assumes that the data set is not linear. This approach assumes that the data points fall perfectly on a line or curve in the absence of random error. Whether the best-fitted curve is linear or not does not affect the ability to interpolate between the experimental points.

In essence, the polynomial method evaluates for nonlinearity: this is the reason for using polynomials. The method consists of two parts. The first part examines whether a nonlinear polynomial fits the data better than a linear one. The second part, performed in cases when a nonlinear polynomial fits the data better then a linear one, assesses whether the difference between the best-fitting nonlinear and linear polynomial is less than the amount of allowable bias for the method, which should be predefined.

5.3.2 Polynomial Regression

Evaluating linearity requires at least five solutions of different concentrations run at least in duplicate. One must know either the analyte concentration or the relationship between the solutions. The relationship between solutions can consist of equidistant intervals or intervals of unequal length, as long

as one knows the relationship between each point. For example, with five solutions having concentrations ranging from 20 to 100 mg/dL, the equidistant solutions would have concentrations of 40, 60, and 80 mg/dL. Make these solutions by following the procedures in Section 4.7 and Appendix A. You can then use 20, 40, 60, 80, and 100 mg/dL for the x-values, or 1,2,3,4, and 5.

Next, perform the polynomial regression analysis for first-, second- and third-order polynomials. This can be accomplished with most commercial statistical software applications.

Order	**Polynomial**	**Regression df (Rdf)**
First	$y = b_0 + b_1 x$	2
Second	$y = b_0 + b_1 x + b_2 x^2$	3
Third	$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3$	4

The first order model is a straight line. This is the equation for the best-fitting line whether the method is linear or not. The second order model describes a relationship where there is a curved response, with either an increasing trend in the response ("curve up," or increasing recovery at higher levels) or decreasing trend ("curve down," or decreasing recovery at higher levels). The third order model fits situations where the response is changing across levels; "sigmoid" or slanted "s-shaped" responses are fitted by this model, as are models where nonlinearity occurs at the ends of the measuring range.

The regression coefficients are labeled as b_i. In the second-order model, b_2 is the nonlinear coefficient; in the third-order model, b_2 and b_3 are nonlinear coefficients. Obtain the standard error of the slope for each nonlinear coefficient, SE_i (available from the regression program output). The next step is to perform a t-test to test whether the nonlinear coefficients are statistically significant; that is, whether the coefficient is significantly different than zero. The first two coefficients (b_0 and b_1) are not tested because they do not reflect nonlinearity. The test is calculated as follows, for b_2 and b_3:

$$t = \frac{b_i}{SE_i}$$

Calculate the number of degrees of freedom from the formula $df = L \cdot R - Rdf$, where L is the number of different sample preparations or concentrations, R is the number of replicates at each preparation or concentration and Rdf is the number of degrees of freedom consumed by the regression analysis. Rdf is the number of coefficients in the regression model (including b_0), as noted above. (In the above example, the third-order polynomial would have $L = 5$, $R = 2$, $Rdf = 4$, and $df = 5 \cdot 2 - 4 = 6$). Look up the critical value for t (two-sided at $\alpha = .05$) in a t-table (Appendix B), or look up the probability of exceeding the observed t. If none of the nonlinear coefficients, b_2 or b_3, are significant ($p > 0.05$ for all), then the dataset is considered linear and the analysis is complete except for the check for high imprecision (Section 5.4). If any of the nonlinear coefficients, b_2 in the second order model, or b_2 or b_3 in the third order model, are significant, i.e., ($p < 0.05$), then the dataset is nonlinear by this protocol. It is important to note that this is merely a test of statistical significance, and indicates only that nonlinearity has been detected, not that it is large enough to affect patient results. Proceed to Section 5.3.3, which deals with the degree of nonlinearity.

5.3.3 Degree of Nonlinearity

Pick the second or third order (nonlinear) polynomial with the best fit by examining the standard error of the regression ($S_{y.x}$). This statistic is a measure of the difference between the measured results and the model, so the model with the lowest value of $S_{y.x}$ provides the best fit for the data.

Calculate the deviation from linearity (DL) at each concentration as follows:

$$DL_i = p(x_i) - (b_0 + b_1 x_i)$$

where the values of x range from $x_1 \ldots x_S$, and $p(x_i)$ is the value of the best-fitting polynomial at point x_i. Therefore, DL_i is the difference between the second order (quadratic) model and the first order (linear) model at every concentration level, or the difference between the third order (cubic) model and the linear model. This is a measure of the difference between the nonlinear model and the linear model, at each of the concentrations measured. The difference is expressed in analyte units so it can be compared with predefined goals. If goals are expressed as percentages, then the DL_i values can be similarly transformed into percentages by dividing DL_i by the concentration at that value (the concentrations are x_i for known value samples, or the average measured value if the x_i are relative concentrations) and multiplying by 100%. The DL_i are calculated only at the sample levels, not any values of the polynomials in between these points (even the presumed nonlinearity might be greatest at levels between the measured levels).

Examine the DL_i at each level and compare with the stated criteria for error at each level. If every DL_i is less than the criterion, then although statistically significant nonlinearity has been detected, it is not important since the amount of nonlinear error is within the goal. If any DL_i exceeds the criterion, there is a possible problem with nonlinearity at that level. There are two approaches to this problem:

(1) Try to find the reason for nonlinearity (sample preparation, interference, instrument calibration, etc.) and remove it. You may wish to contact the representative for your system or instrument for help with this investigation.

(2) Examine the graph of response vs. concentration and determine whether the nonlinearity is at either end of the range of concentrations or in the middle of the range. If the nonlinear concentration is at either end, one option is to remove the point where DL_i was too large, and re-run the statistical analysis. *This will of course reduce the linear range.*

5.4 Considerations for Random Error

The assessment of linearity so far has not considered the contribution made by random error (in this protocol, estimated as repeatability). Random error results from the random variability, (variability in the analytical system) and can lead to poor ability to detect nonlinearity. The best estimate of repeatability is the pooled difference between the *L* sets of replicates. The pooled difference between replicates is an overall average measure of variability that is independent of analyte level. This is the "repeatability" of the method, and is noted as SD_r (or CV_r for proportional error). If the estimates of SD_r are reasonably equal across levels, then repeatability is constant across levels (constant SD_r). If the difference is much larger at high concentrations, then repeatability may be approximately proportional to the reference concentration (constant CV_r). If repeatability is proportional to the concentration, then error calculations should be made on percentage differences rather than absolute differences. Repeatability can be calculated by analysis of variance, as the square root of the mean square for error.

It can also be calculated easily by hand for two replicates using the differences between replicates, as follows:

- calculate the difference between two replicates at each level;
- square the differences between the replicates;
- sum the squared differences;
- divide by the number of levels (L) x 2;
- take the square root.

The formula is as follows (for two replicates and L levels):

$$SD_r = \sqrt{\frac{\sum_{i=1}^{L}[r_{i1} - r_{i2}]^2}{2 \times L}}$$

r_{i1} and r_{i2} can be either the actual results of the procedure, or they can be expressed as percentages of the solution mean, (although the same units must be used for all levels). If percentage differences are used, then the result is a CV_r, not SD_r.

For situations with more than two replicates, the estimate of random error should come from an analysis of variance, which is available on most spreadsheet programs. Following is the calculation formula:

$$SD_r = \sqrt{\frac{\sum_{i=1}^{L}\sum_{j=1}^{R}[r_{ij} - r_{i.}]^2}{L \times (R-1)}}$$

where R = number of replicates at each level (j=1,…,R);
L = number of levels (i =1,…,L);
r_i = average result at level i.

Compare SD_r with the goal for repeatability, either as is, or expressed as a percentage of any concentration. If SD_r is larger than the goal, then precision may not be adequate for a reliable determination of linearity. That is, the means of the replicates at each level may be too uncertain for a true determination of the linear relationship. In this case, the analyst may wish to examine the instrument or method performance to discover the reason for imprecision, correct the problem, and then re-run the experiment. If the method's performance is consistent with past repeatability estimates, then the analyst could simply double the number of replicates. This would lower the standard error of the means by about 40%.

In some cases, where repeatability is severely different across concentrations, it may be preferable to use weighted regression for the first-order (linear) model. The weights to be used are the inverse of the variance of the replicates at each level. This requires more sophisticated statistical software and perhaps the assistance of a statistician.

The linearity protocol is illustrated in the figure below. Each task in the flowchart has a parenthetical note referring to the appropriate section number in the text.

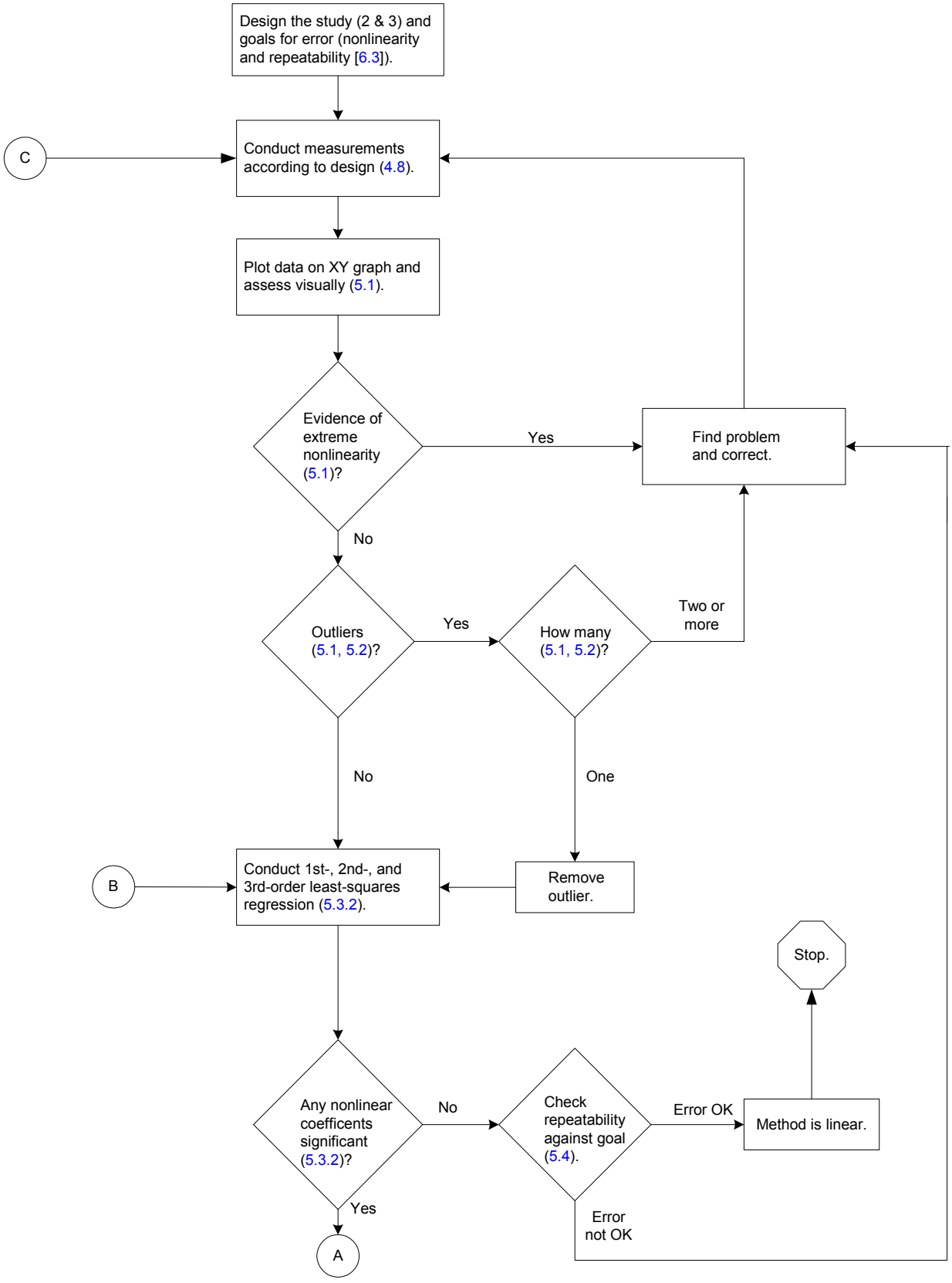

Figure 1. Flowchart for EP6 Linearity Procedure (References to appropriate sections in the text are in parenthesis.)

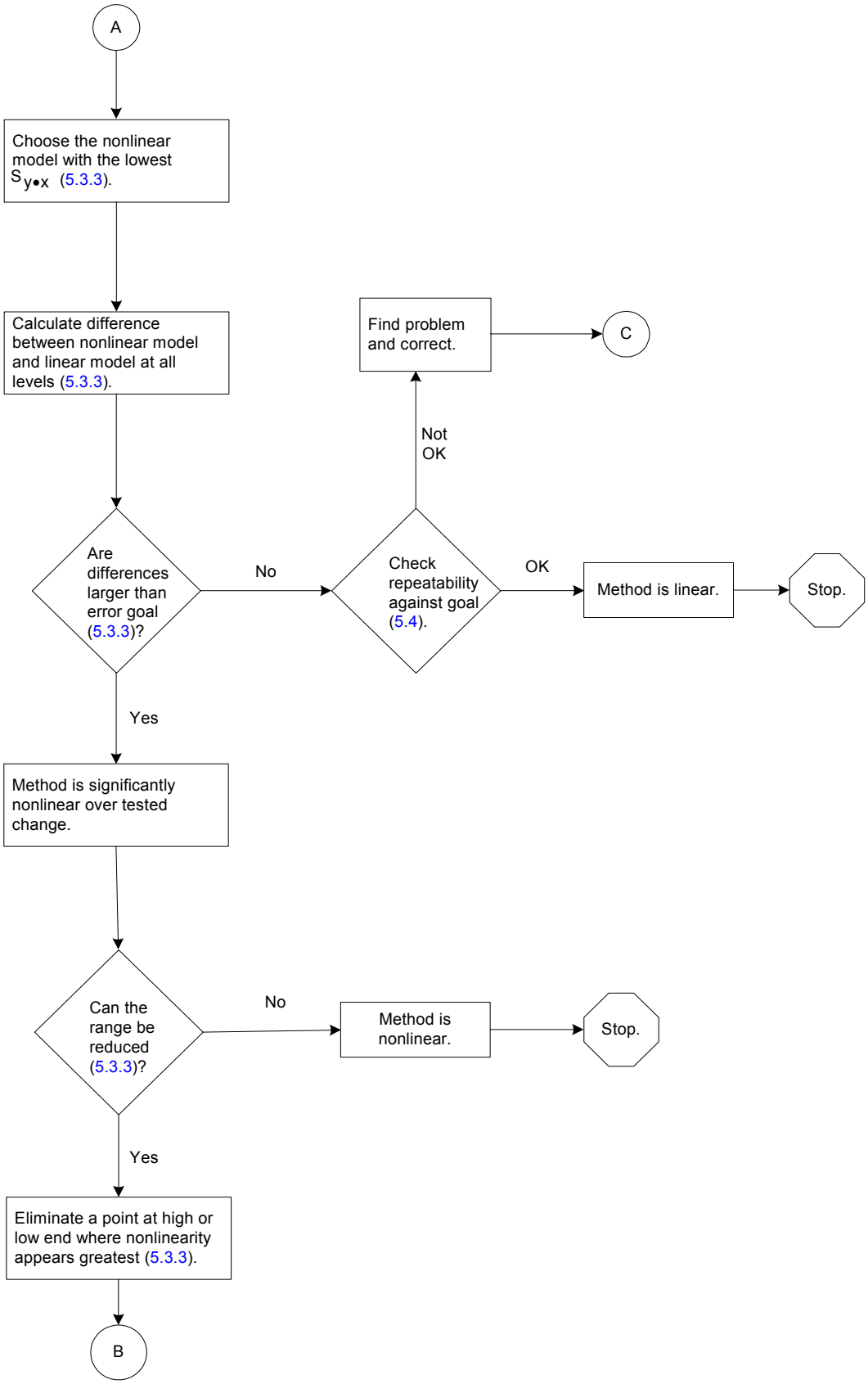

Figure 1. (Continued)

6 Linear Range Claims

6.1 Graphical Approaches

It is important that the statistical findings be compared with a graphical representation of the analytical results (Y) versus the concentrations (X). A simple graph shows the extent to which the results follow a straight line and the areas where nonlinearity is most severe. The graph can be useful for overall understanding and possible rejection, but not acceptance, of a linearity claim, nor determination of the magnitude of nonlinearity at every level. The visual plot alone is insufficient to establish or check claims.

The magnitude of nonlinearity is best shown in a "Difference Plot" where the differences between the first- and third-order models is plotted on the vertical axis and the concentration at each point is shown on the horizontal axis. The allowances for error may also be added to this graph, with any percentage differences transposed into actual units. Examples are shown in Appendix C.

6.2 Verification of Claims for Linear Range

The procedures used in this guideline should help users to demonstrate that their systems meet the linear range claimed for the method. Manufacturers and laboratories that develop methods may establish claims in many different ways, including the use of procedures given in this guideline. No matter how the claims are established, they can be verified with this protocol.

6.3 Allowable Error and Goal Setting

The laboratory must determine its own goals for measurement error for every analyte (see the current version of NCCLS document EP21—*Total Error for Clinical Laboratory Methods*). The goals should be based on the needs of the clients of that laboratory and an understanding of the capabilities of the method as used in that laboratory. Requirements for measurement error can be used to determine goals for the various sources of measurement error, such as various components of precision, trueness, and nonlinearity. Various models have been proposed for determining goals for measurement error.[15-23] It is beyond the scope of this document to recommend any procedures for setting goals.

However, following are considerations for goals for linearity:

- Goals for linearity should be derived from goals for bias, and should be less than or equal to those goals.

- Goals for bias should be less than or equal to goals for measurement error.

When the concentrations of the test samples are unknown, the goals for linearity should be in relative (percentage) units.

6.4 Effects of Other Errors

An unacceptable experimental outcome can be due to procedural error; instrument, reagent, or calibration problems; environmental incompatibility; methodological error; or matrix incompatibility, which includes interferences. Each of these can result in poor repeatability or large bias. The first step in determining the problem's source is to analyze the data and the method and try to understand the reason for the problem. It may be advisable to repeat the experiment, or to increase the number of concentrations (especially at concentrations that demonstrate unacceptable curvature, or increase the number of replicates, especially if excessive imprecision is apparent. If repeatability is worse than expected, there may be a reason to stop the experiment until a problem is identified and corrected. The committee encourages users to contact the

manufacturer to help resolve any problems that remain unresolved. The user may find the following protocols useful if unacceptable problems are found:

- Technical error such as interference (see the most current edition of NCCLS document EP7— *Interference Testing in Clinical Chemistry*).
- Sample carryover and drift in test values as well as preliminary evaluation of imprecision and bias (see the most current edition of NCCLS document EP10— *Preliminary Evaluation of Quantitative Clinical Laboratory Methods*).
- Analytical imprecision (see the most current edition of NCCLS document EP5— *Evaluation of Precision Performance of Clinical Chemistry Devices*).
- Bias (see the most current edition of NCCLS document EP9— *Method Comparison and Bias Estimation Using Patient Samples*).

Excessive unexpected errors should be evaluated to determine the usability of the method or instrument for this purpose. Measurement errors due to imprecision, poor trueness, carryover, and drift can affect the apparent linear range: these errors can result in erroneous interpretation of the data.[13]

6.5 Assumptions

The assumptions used in this protocol for the statistical evaluation of linear range include the following:

- The sample levels are known without error (actual levels known, or known relative to each other).
- The linear range is tested only to the lowest and highest concentrations that demonstrate acceptable performance.
- The linear range is evaluated for the system's final output (concentration or activity) and not for the instrument signal, which may be further processed by the system.
- The samples used are free from interferences that would invalidate the experiment.
- The testing system's operation meets its other performance claims within the tested linear range.
- Tests for the significance of the regression coefficients also assume that replicates are normally distributed at each level, and the variance of this distribution is constant across all levels.

6.6 Guidelines for Stating a Claim for the Linear Range

6.6.1 Suggested Wording of Claims

Suggested wording of the claim is as follows unless the method developer provides alternative wording that includes the range of linearity and maximum difference. If any claim is made for conformance with this linearity protocol, it must include the limits of acceptability used for assessing linearity. Fill in the correct words or values for items in parentheses:

*For **(analyte)** by **(method)**, the method has been demonstrated to be linear from **(lower limit)** to **(upper limit)**, within **(goal or measured max diff)** in this interval.*

6.6.2 Optional Additional Information

It may be useful to add the allowable difference (error) boundary at clinical decision points. If difference (error) limits are not stated at clinical decision points, the user can assume that the analytical difference (error) boundaries are proportional throughout the range.

References

[1] Kroll MH, Emancipator K. A theoretical evaluation of linearity. *Clin Chem.* 1993; 39:405-413.

[2] Emancipator K, Kroll MH. A quantitative measure of nonlinearity. *Clin Chem.* 1993; 39:766-772.

[3] Kroll MH, Emancipator K, Floering D, Tholen D. An algorithm for finding the linear region in a nonlinear data set. *Computers in Biology and Medicine.* 1999; 29:289-301.

[4] Kroll MH, Praestgard J, Michaliszyn E, Styer P. Evaluation of the extent of nonlinearity in reportable range studies. *Arch Pathol Lab Med.* 2000; 124:1331-1338.

[5] Krouwer JS, Schlain B. A method to quantify deviations from assay linearity. *Clin Chem.* 1993; 39:1689-1693.

[6] Tholen DW. Alternative statistical techniques to evaluate linearity. *Arch Pathol Lab Med.* 1992; 746-756.

[7] International Organization for Standardization. In Vitro diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials. ISO/FDIS 17511. Geneva:2002.

[8] International Organization for Standardization. Statistics – Vocabulary and symbols – Part 1: Probability and general statistical terms. ISO-3534-1. Geneva: 1993.

[9] International Organization for Standardization. International of Vocabulary of Basic and General Terms of Metrology (VIM). Geneva: 1993.

[10] Draper NR, Smith H. *Applied Regression Analysis*. 3rd ed. New York, NY: John Wiley & Sons, Inc.; 1998.

[11] Kutner MH, Nachtschiem CJ, Wasserman W, Neter J. *Applied Linear Statistical Models.* 4th ed. New York, NY: McGraw Hill Book Co.; 1996.

[12] Crow EL, Davis FA, Maxfield MW. *Statistics Manual*. Dover Publications; 1960.

[13] Daniel C. Calibration designs for machines with carryover and drift. *J. Qual. Technol.* 1975;103-108.

[14] ASTM E-178-94. Standard Practice for Dealing with Outlying Observations. West Conshohocken, PA: American Society for Testing and Materials, 1994.

[15] Tonks D. A Quality control program for quantitative clinical chemistry evaluations. *Can J Med Technol.* 1968;30:38-54.

[16] Ehrmeyer SS, Laessig RL, et al. 1990 Medicare / CLIA final rules for proficiency testing: Minimum intralaboratory performance characteristics (CV and bias) need to pass. *Clin Chem.*1990;36:1736-1740.

[17] Westgard JO, Burnett RW. Precision requirements for cost-effective operation of analytical processes. Clin Chem. 1990; 36:1629-1632.

[18] Garber CC, Carey RN. Evaluation of Methods. In Kaplan LA and Pesce AJ, eds. *Clinical Chemistry: Theory, Analysis, and Correlation*. 3rd ed. Mosby: 1996;402-423

[19] Fraser CG. The application of theoretical goals based on biological variation data in proficiency testing. *Arch Pathol Lab Med*. 1988; 112:404-415.

[20] Fraser CG. Biological variation in clinical chemistry: An update(collated data). 1998 – 1991. *Arch Pathol Lab Med*. 1992;116:916-923.

[21] Fraser CG, Hyltoft Peterson P, Ricos C, Haeckel R. Proposed quality specifications for imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Clin Biochem*. 1992;30:311-317.

[22] Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Lab Sci*. 1989;27: 409-37.

[23] Ricos C, Alvarez V, Cava F, et al. Current data bases on biological variation: pros, cons, and progress. *Scan J Clin Lab Invest*. 1999; 59: 491-500.

Appendix A. Dilution Schemes

This appendix gives dilution schemes for samples with equally spaced concentrations. The actual concentrations are not important for the purposes of assessing linearity. However, the linear range will be defined by the highest and lowest measured concentrations where the response is linear. In all schemes, start with a high concentration sample and a low concentration sample whose concentrations meet or exceed the range of interest, and with sufficient volume to allow replicate assays of all points, when all samples are prepared. For example, if 0.5 mL is required for each assay, and duplicate assays are to be used for each of five points, then a total of 5 mL is needed, or at least 2.5 mL in each of the high and low concentration samples. In all dilution schemes, use fractional parts of each sample.

S=5 samples

1: Low (L)
2: 0.75L+0.25H
3: 0.50L+0.50H
4: 0.25L+0.75H
5: High (H)

S=6 samples

1: Low (L)
2: 0.8L+0.2H
3: 0.6L+0.4H
4: 0.4L+0.6H
5: 0.2L+0.8H
6: High (H)

S=7 samples

1: Low (L)
2: 0.833L+0.167H
3: 0.667L+0.333H
4: 0.500L+0.500H
5: 0.333L+0.667H
6: 0.167L+0.833H
7: High (H)

S=8 samples

1: Low (L)
2: 0.857L+0.143H
3: 0.714L+0.286H
4: 0.571L+0.429H
5: 0.429L+0.571H
6: 0.286L+0.714H
7: 0.143L+0.857H
8: High (H)

S=9 samples

1: Low (L)
2: 0.875L+0.125H
3: 0.750L+0.250H
4: 0.625L+0.375H
5: 0.500L+0.500H
6: 0.375L+0.625H
7: 0.250L+0.750H
8: 0.125L+0.875H
9: High (H)

S=10 samples

1: Low (L)
2: 0.889L+0.111H
3: 0.778L+0.222H
4: 0.667L+0.333H
5: 0.556L+0.444H
6: 0.444H+0.556H
7: 0.333L+0.667H
8: 0.222L+0.778H
9: 0.111L+0.889H
10: High (H)

S=11 samples

1: Low (L)
2: 0.9L+0.1H
3: 0.8L+0.2H
4: 0.7L+0.3H
5: 0.6L+0.4H
6: 0.5L+0.5H
7: 0.4H+0.6H
8: 0.3L+0.7H
9: 0.2L+0.8H
10:0.1L+0.9H
11:High (H)

Appendix B. t table

Following is a limited table for the "Students t" distribution. Shown are two-sided 90, 95, and 99 percentiles for various degrees of freedom. These values apply equally to positive and negative statistics. For example, the central 95th two-sided percentile value for 10 degrees of freedom is 2.228. So t statistics of +2.23 (or larger) and –2.23 (or smaller) are statistically significant at the $\alpha = .05$ level.

DF	.90	.95	.99
6	1.943	2.447	3.707
7	1.895	2.365	3.500
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.794	2.201	3.106
12	1.782	2.179	3.054
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.132	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
21	1.721	2.080	2.831
22	1.717	2.074	2.819
23	1.714	2.069	2.807
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750

Appendix C. Examples

Example 1. IgM

A laboratory prepares five equally spaced concentrations for duplicate analysis of IgM. Assume the laboratory has set default criteria of 2% for repeatability and 5% for nonlinearity.

The laboratory conducts the ten analyses in random order in a single run. The results are recorded and they are visually reviewed for approximate linearity and agreement between replicates. This is to check for outliers and whether or not the results make sense. The repeatability estimate is calculated and compared with the goal. Regression analyses then are used to check for nonlinear response patterns. The differences between linear and nonlinear models are assessed, and the differences are compared with a preset criterion.

Following are results of the linearity study:

Dilution	Replicate #1	Replicate #2	Mean
1	26.5	26.2	26.35
2	139	138	138.5
3	269	273	271.0
4	337	343	340.0
5	409	404	406.5

There are no apparent outliers; but there is an indication of repeatability differences at different levels, with very small differences between replicates at the lower levels and larger differences at the higher levels. The differences between replicates are examined as follows:

Dilution	Mean	Difference	Diff ^{2}/2	% Diff	%Diff ^{2}/2
1	26.35	0.3	.045	1.14	.648
2	138.5	1.0	.5	0.72	.261
3	271.0	4.0	8	1.48	1.089
4	341.0	6.0	18	1.76	1.548
5	406.5	5.0	12.5	1.23	.756

The sum of squared differences ($D^2/2$) is 39.05, and the average of that sum is 7.81; the square root of that is 2.8, which is the pooled repeatability estimate. However, the differences between replicates appear to be proportional rather than constant, so the overall pooled difference is not valid. In the last two columns the percentage differences are used, which calculate a pooled difference SD of 0.9%.

In reviewing the above, an estimate of 0.9% repeatability seems more reasonable than 2.8. With random error at SD=0.9%, analytic error will usually be less than 2%, as required by the laboratory's default criteria, but it could exceed 3.6% (4SD), and so violate the error criterion. A more specific estimate of random error should be performed with procedures recommended in NCCLS documents EP5– *Evaluation of Precision Performance of Clinical Chemistry Devices,* or EP15– *User Demonstration of Performance for Precision and Accuracy,* where the estimates are derived at individual concentrations, in contrast to this one where the SD was determined across five different concentrations.

The following table shows the results from regression analysis. Only the regression coefficients are shown, along with their Standard Errors and the t statistic. The standard error of regression ($S_{y.x}$) is also shown, as a measure of fit. This shows that the second-order model fits much better than the first order, and that there is a statistically significant coefficient for the second-order term (t=5.7 exceeds the tabled limit of 2.365). The third-order model has the same fit, and no significant coefficients. Note that we are interested only in the coefficient for the nonlinear terms, not the significance of the intercept or first-order term.

Results of Regression Analysis

Order	Coef. Symbol	Coefficient Value	Coefficient SE	t-test	Degrees freedom	Std Error Regression
First	b_0	-52.07	16.92	3.1		
First	b_1	96.18	5.10	18.8	8	22.8
Second	b_0	-129.47	15.62	8.3		
Second	b_1	162.52	11.91	13.6		
Second	**b_2**	-11.06	1.95	**5.7 (*)**	7	**10.3**
Third	b_0	-97.48	35.88	-2.7		
Third	b_1	117.58	46.91	2.5		
Third	**b_2**	6.08	17.41	0.3		
Third	**b_3**	-1.90	1.92	-1.0	6	**10.3**

A comparison of the models shows very large differences at the low and middle concentrations, and better agreement at the high concentrations. These differences exceed the laboratory's goals for error at the low, middle, and high concentrations, and so the method has unacceptable nonlinearity.

Actual Mean	Predicted 1^{st}-order	Predicted 2^{nd}-order	Difference 2^{nd}-1^{st}	% Difference
26.4	44.1	22.0	**-22.1**	**-50.2**
138.5	140.3	151.3	**11.0**	**7.8**
271.0	236.5	258.6	**22.1**	**9.3**
340.0	332.7	343.7	11.0	3.3
406.5	428.8	406.7	**-22.1**	**-5.2**

Comments

This case is an example where the third-order regression has no significant nonlinear terms, while the second-order fit has a significant coefficient for the nonlinear term. The second- and first-order models are shown in Figure C1, and the nonlinearity is visually apparent.

Note that at four concentrations the percentage difference between first- and second-order fits exceeds the goal of five percent. The system is judged nonlinear and will be restudied. This is shown in Figure C2 (Difference Plot). The largest percentage differences are at the lowest concentration, but nonlinearity exists throughout the range.

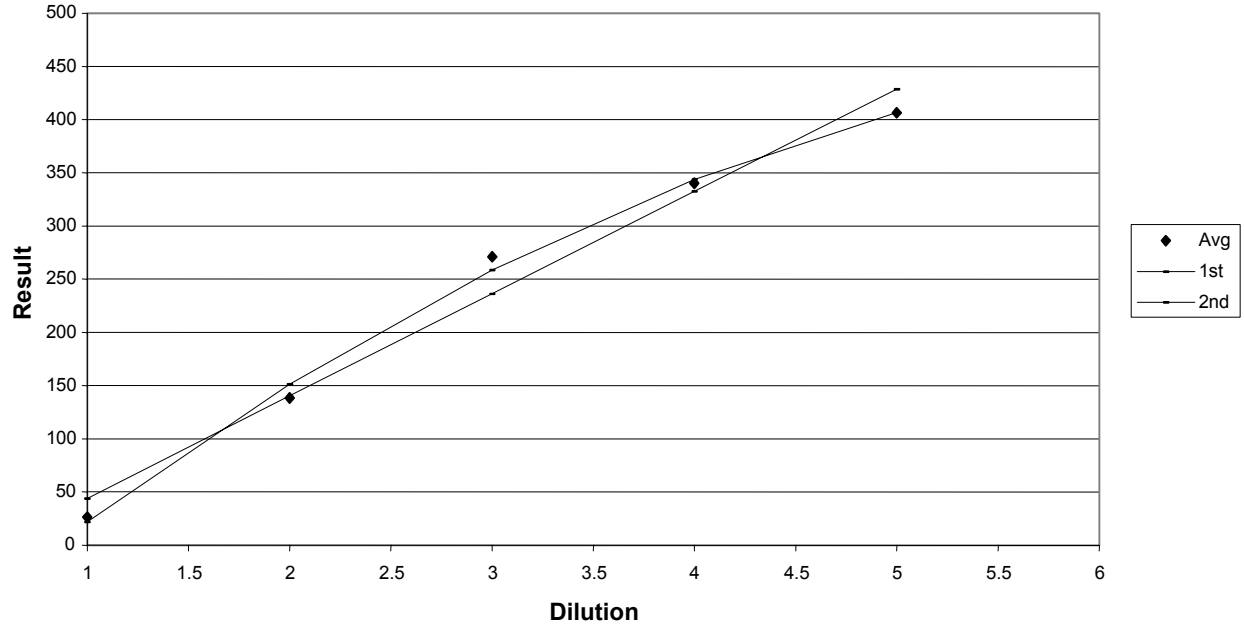

Figure C1. IgM Linearity Study—Dilutions 1-5

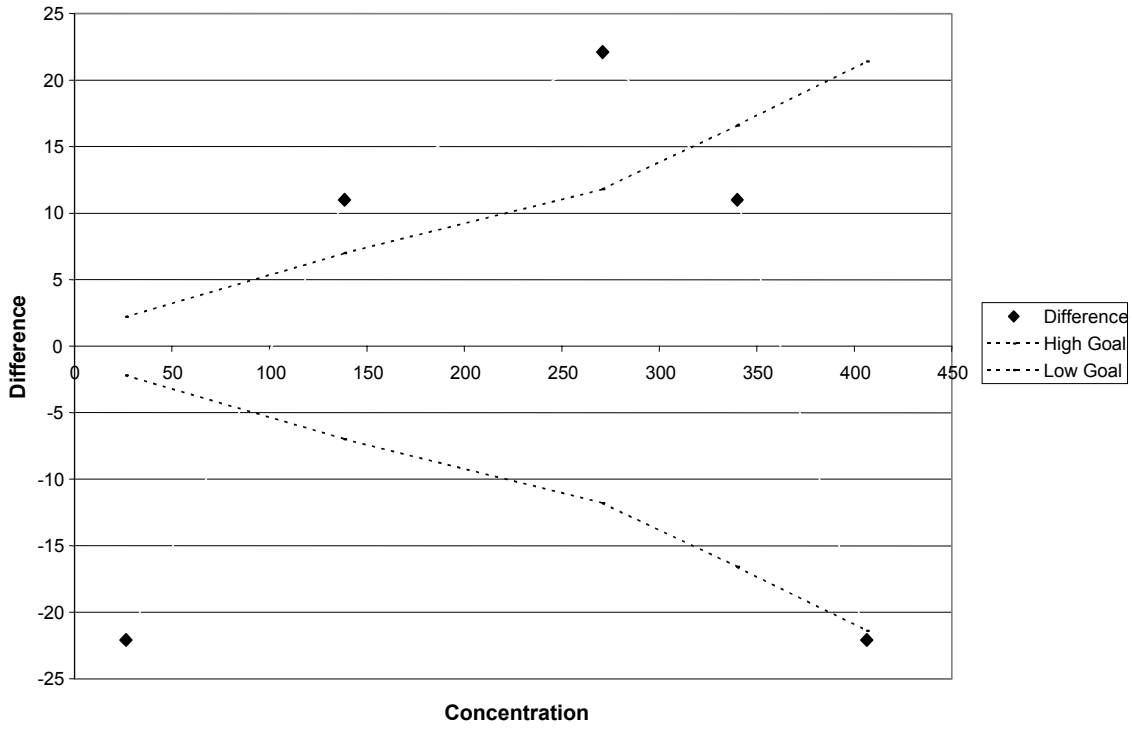

Figure C2. IgM Linearity Difference Plot

Example 2. Calcium

Assume the laboratory has set error criteria of 0.20 mg/dL for repeatability and 0.20 mg/dL for nonlinearity. These are not percentage criteria because error for Calcium tends to be constant rather than proportional. As with the first example, the laboratory prepares the samples and conducts duplicate assays at all six levels.

The results are placed in the table below and simple calculations are used to calculate the means and replicate differences. The pooled repeatability estimate is also determined by simple calculations. Regression analyses are then used to check the nonlinear models. There is a significant higher-order model, so the linear and nonlinear models are compared to estimate the size of the nonlinear error.

Data Table

Dilution	Rep1	Rep2	Mean
1	4.7	4.6	4.65
2	7.8	7.6	7.70
3	10.4	10.2	10.30
4	13.0	13.1	13.05
5	15.5	15.3	15.40
6	16.3	16.1	16.20

There are no apparent outliers.

Dilution	Mean	Difference	$D^2/2$
1	4.65	0.1	.005
2	7.70	0.2	.02
3	10.30	0.2	.02
4	13.05	0.1	.005
5	15.40	0.2	.02
6	16.20	0.2	.02

The pooled replicate difference SD is 0.12 (repeatability), which ranges from 0.6% to 2.6%. It appears to be constant across the levels.

The means and regression lines (first and third only) are shown in Figure C3, below.

Regression analysis shows that the nonlinear term in the second-order model is highly significant, but the third-order nonlinear coefficients are also significant (t = 2.6 and -3.8). These statistics exceed the criterion of 2.306 for 8 degrees of freedom. The third-order model also has much lower Standard Error, indicating a better fit than the second- or first-order models.

Order	Coefficient	Value	SE	t-test	Std err $S_{y.x}$	Degrees freedom
1st	b_0	2.86	0.44	6.5		
1st	b_1	2.39	0.11	21.2	.667	10
2nd	b_0	0.82	0.40	2.1		
2nd	b_1	3.92	0.26	15.2		
2nd	**b_2**	-0.22	0.04	**-6.0**	.313	9
3rd	b_0	2.48	0.50	4.9		
3rd	b_1	1.82	0.57	3.2		
3rd	**b_2**	0.48	0.18	**2.6**		
3rd	**b_3**	-0.07	0.02	**-3.8**	.197	8

The first- and third-order models are "solved" for the coded concentrations 1-6. The differences range from 0.9 to 0.1, or 12% to 0.5%. Five points have differences that exceed the laboratory's criterion for nonlinear error (0.20).

Result Means	Predicted 1st order	Predicted 3rd order	Difference	% Difference
4.65	5.25	4.71	-0.54	-11.6
7.70	7.63	7.50	-0.13	-1.7
10.30	10.02	10.45	0.43	4.2
13.05	12.41	13.15	0.74	5.6
15.40	14.80	15.22	0.42	2.7
16.20	17.19	16.26	-0.93	-5.7

It is decided to eliminate the last point. If the other five points are acceptably linear, then the method has a range of around 4.5 to 15, which suits the laboratory's needs. The revised analysis is shown below.

Order	Coefficient	Value	SE	t-test	Std err $S_{y.x}$	Degrees freedom
1st	b_0	2.16	0.15	14.3		
1st	b_1	2.68	0.05	59.0	.204	8
2nd	b_0	1.54	0.19	8.2		
2nd	b_1	3.22	0.14	22.4		
2nd	**b_2**	-0.09	0.02	**-3.8**	.124	7
3rd	b_0	1.47	0.47	3.15		
3rd	b_1	3.32	0.61	5.45		
3rd	**b_2**	-0.13	0.23	-0.56		
3rd	**b_3**	0.004	0.02	0.17	.134	6

In the revised analysis, the second-order term in the second-degree polynomial is significant (t = -3.8 exceeds critical value of 2.365 for 7 degrees of freedom) and the nonlinear terms in the third-order model are not significant. The standard errors also show that the second-order model fits better than the first- or third-order models. Differences in predicted values between the first- and second-order models are shown below. None of these differences exceeds the laboratory criterion of 0.2 mg/dl, so the method is acceptably linear over the limited range. The reduced model is shown in Figure C3, showing the five points and the first- and second-order models.

Result Means	Predicted 1st order	Predicted 2nd order	Difference	% Difference
4.65	4.85	4.67	-0.18	-3.9
7.70	7.54	7.62	0.08	1.0
10.30	10.22	10.40	0.18	1.8
13.05	12.90	12.99	0.09	0.7
15.40	15.59	15.41	-0.18	-1.2

Comments

In this case, dropping one point and rerunning the statistical analysis shows a nonlinear response, but the differences are smaller than the laboratory criterion. The method is assumed to be acceptably linear over the reduced range.

Figures C3 and C4 show the first- and third-order models, and the differences between them. Five differences are outside the limit of 0.2. Figures C5 and C6 show the models with the highest point removed, and the differences between the first- and second-order models. The differences are all within the goals, even though there remains visual nonlinearity.

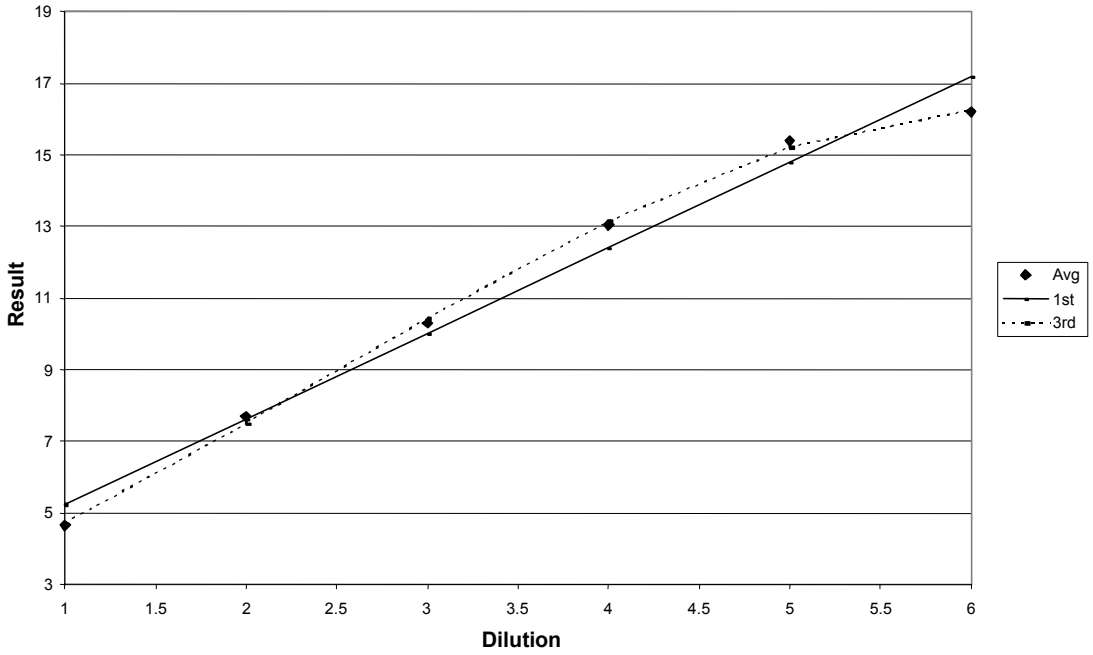

Figure C3. Calcium Linearity Study—Dilutions 1-6

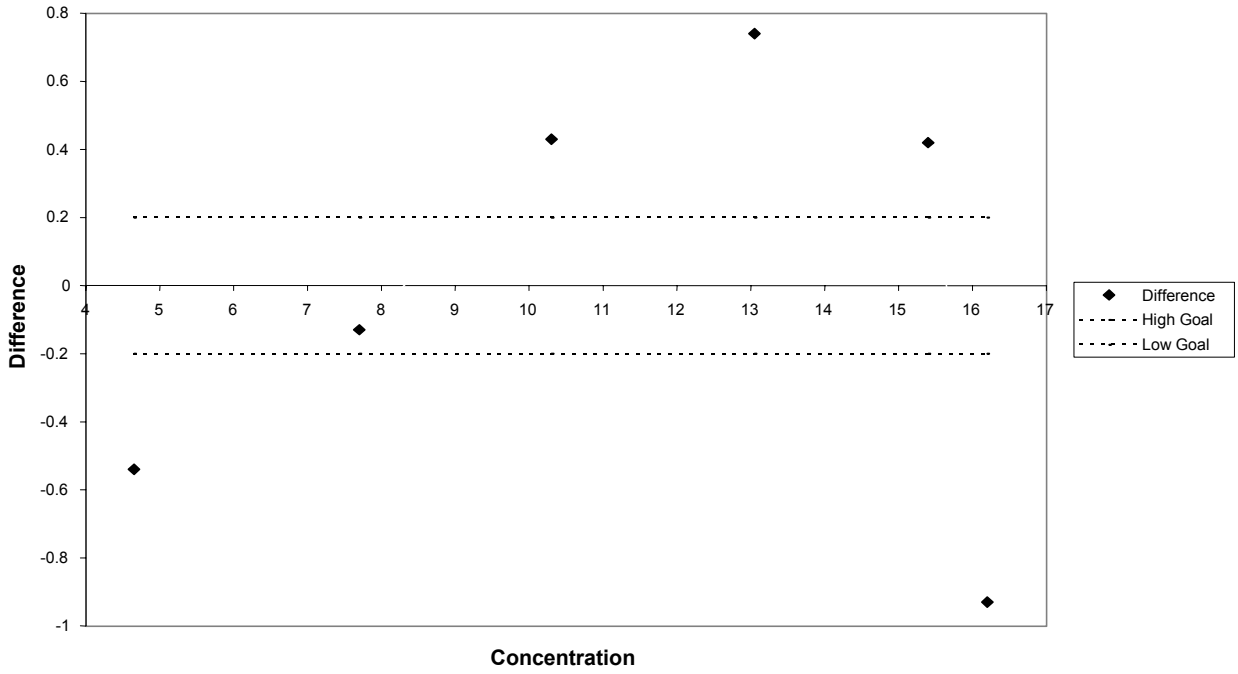

Figure C4. Calcium Difference Plot—Dilutions 1-6

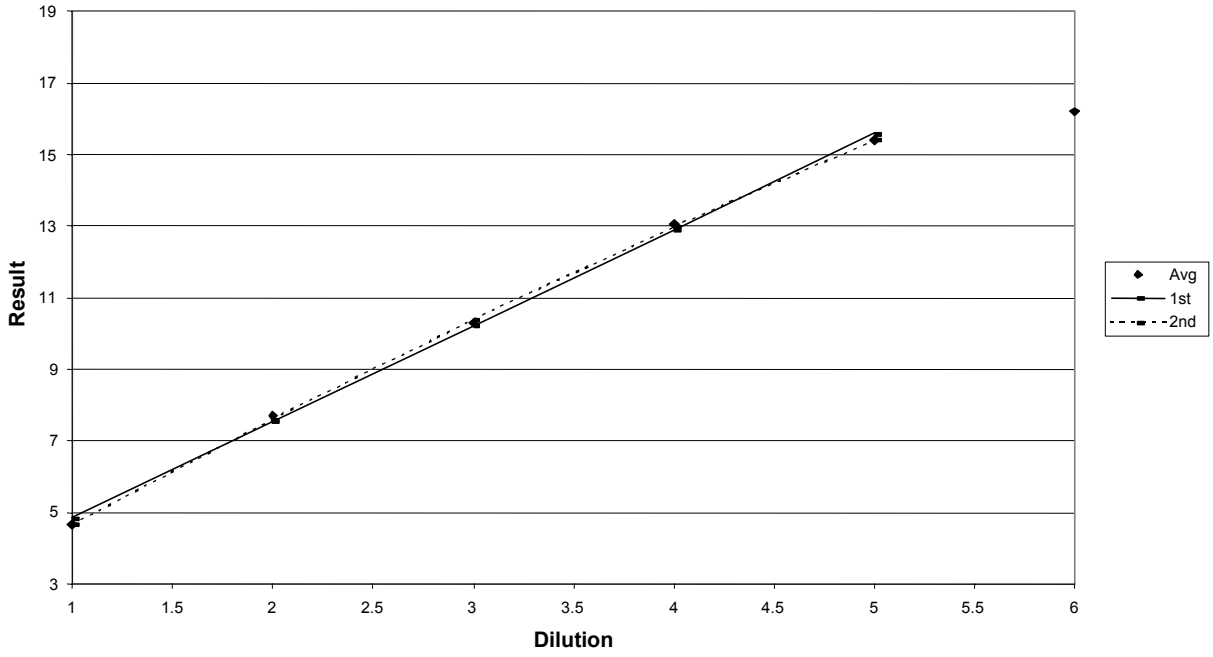

Figure C5. Calcium Linearity Study—Dilutions 1-5

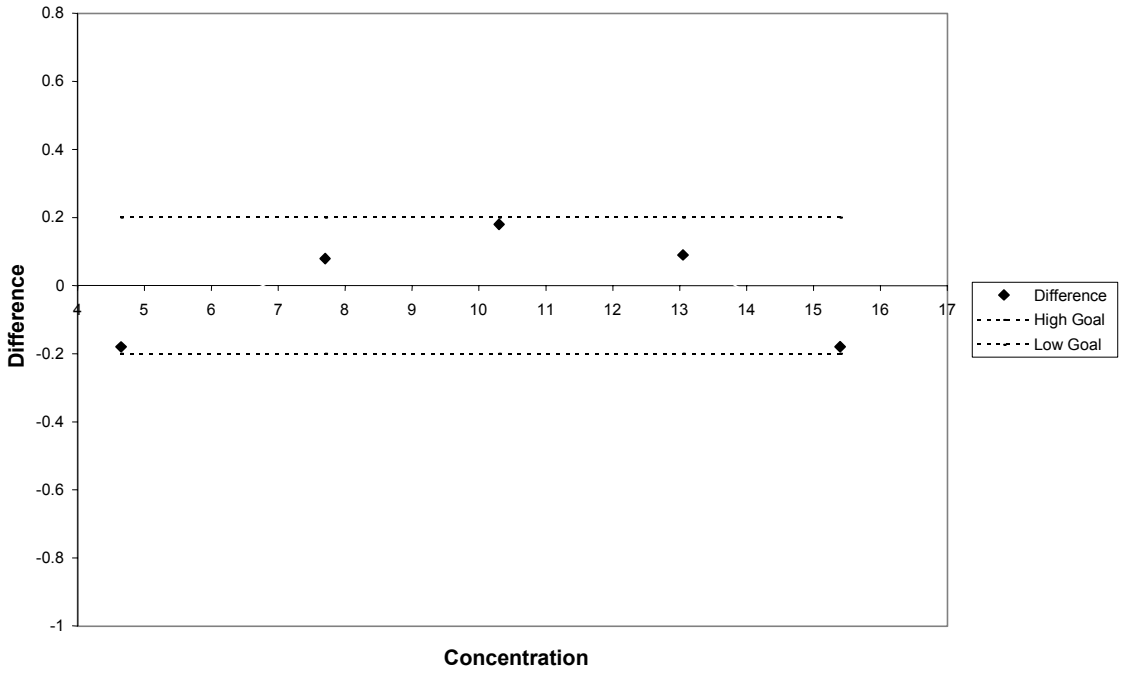

Figure C6. Calcium Difference Plot—Dilutions 1-5

NCCLS consensus procedures include an appeals process that is described in detail in Section 9 of the Administrative Procedures. For further information, contact the Executive Offices or visit our website at www.nccls.org.

Summary of Comments and Working Group Responses

EP6-P2: *Evaluation of the Linearity of Quantitative Analytical Methods; Proposed Guideline— Second Edition*

General

1. A valuable guide and document.

- **The working group appreciates the comment.**

2. I am very concerned about the utility of this procedure. The statistical logic is not adequately explained for use by anyone but a statistician. I am comfortable with statistics and, as a manufacturer, use the parameters routinely; however, I did not find the explanations adequate for my understanding.

- **Linearity checks with statistical techniques require some understanding of common statistical methods. These techniques are explained in common statistical texts. (See references 7-9). In addition, the working group has recommended that software be provided to complement this guideline and assist with the statistical calculations. Also, to make the linearity assessment protocol easier to follow, a flowchart has been inserted at the end of Section 5.4.**

3. The document makes several statements about statistical analysis being the only acceptable means of evaluating or establishing linearity. My company does neither, and we have had little or no issues with our methods of evaluation. In general, it is unacceptable for a guideline to be prescriptive to users about the application of any protocol. This *must* be changed.

- **The title of the document has been changed to *Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline:* to alert the user that the protocol requires familiarity with common statistical methods; and to reflect that the protocol in the document is one valid, objective approach to assess linearity, but not the only approach. There is abundant evidence that this particular technique is a valid procedure for evaluating linearity; the CAP Linearity Surveys have used a similar protocol successfully for a wide variety of analytes in chemistry, toxicology, hematology, and immunology, and for all common methods in these fields.**

4. In general, this document presents a method for calculation of linearity that is too difficult for the average laboratorian to perform. We are very concerned that if this document is approved, laboratories all over the U.S. will be cited by their inspectors for failure to perform their linearity studies following this guideline. Specifically, the method is too statistical. It is not necessary for laboratories to be able to verify the linearity of their methods with such statistical accuracy.

- **See the responses to Comments 2 and 3.**

5. I have read through the document and the comments. I agree that this not something I would want to hand to a laboratory technician to accomplish. This requires sophisticated software and a lot of thought. There are not enough explanations and examples. Most NCCLS protocols have worksheets for doing the analysis by hand with the aid of a calculator. This would require a computer, with a statistics program to calculate the cubic fit of the data.

   The only suggestion I would have would be to have a procedural list that someone could refer to:

   1. Preparation of samples –Section 3.3 and Section ?
   2. Replicate guidelines – Assay variability vs. sample size.
   3. Etc.

- **For the first part of this comment, see responses to Comments 2 and 3. For the second part, the working group cannot respond without further information.**

6. Users should be performing method comparison studies using native samples that incorporate sources of error. Goals should be set for acceptable levels of total error. If, after performing the method comparison, it is seen that the total error is too high, then the user can dig deeper into the causes of that error. Perhaps they can then adjust their analyzer or their assay to achieve better precision, or less bias. If the total error appears to increase at the ends of the reportable range, then the user should suspect nonlinearity, and perform linearity testing. In this case, a simple visual graph of the results is sufficient to determine if nonlinearity is the cause for the increased total error.

- **This comment gives important considerations that should be addressed in a document on setting goals for error. However, the current document states that a visual assessment is suitable to determine significant nonlinearity.**

7. A statistical determination of linearity is almost never needed by the clinical laboratories. Rather, it is important to know if clinically relevant results are being produced. Laboratorians should ask: "is the total error of my test within allowable clinical limits."

- **The title has been changed to clearly indicate that this document provides a statistical approach to evaluation of linearity. Please refer to Section 2.1 for situations where this procedure is appropriate.**

8. If the authors are intending this document to be used by manufacturers, then it is too prescriptive. Manufacturers are free to determine their own methods for linearity. My company uses various approaches, including a similar approach as described here. However, our approach is somewhat simpler in that it does not require a stepwise recalculation of the results. Yet, we would never expect the average laboratory to be able to repeat our study, as it too is far more statistical than is necessary for the lab. Our recommendation is that this document either be: scrapped in its entirety, or clarified that this is an example of one method that can be used to establish linearity. It is not the only method. The laboratory should choose a method that is appropriate for their needs.

- **Please see the response to Comment 7.**

9. We have serious reservations of the utility of this guideline in its current form. Labs should be instructed to focus on total error across the reportable range. Only when total error is unacceptable, should they then be concerned with the components of the error. Labs then need a very simple method to quickly assess if nonlinearity is the cause of the unacceptable total error.

- **The following bullets were added at the end of Section 6.3 to further define considerations for setting goals for measurement error, bias, and nonlinearity:**

  - **Goals for linearity should be derived from goals for bias, and should be less than or equal to those goals.**

  - **Goals for bias should be less than or equal to goals for measurement error.**

  - **When the concentrations of the test samples are unknown, the goals for linearity should be in relative (percentage) units.**

- **Also, please see the response to Comments 98, 99, and 100.**

10. It is almost never necessary for labs to establish the statistical linearity of their assay. Verification of a manufacturer's linearity claim can be performed with visual inspection of the plotted results, keeping in mind the clinically allowable error.

- **See the response to Comment 7.**

11. The statistical approach as outlined seems very complex for the normal moderately complex lab to perform. Simplify. Visual evaluation is in most cases appropriate for laboratories. This document needs to allow for that method as well as statistical means.

- **See the response to Comment 7.**

12. This procedure is too complicated for the laboratory to perform and it is not a requirement of the regulations. Are we expecting the lab to use this procedure to verify the reportable range of the instrument? This document does not serve the target audience.

- **The title has been changed to clearly indicate that this document provides a statistical approach to evaluation of linearity. If a laboratory decides to use a statistical procedure, this is the preferred approach.**

13. In general, this document presents a method for calculation of linearity that is too difficult for the average laboratorian to perform. The level of sophistication may be appropriate for a statistician's assessment of linearity, but is well beyond the scope of the average laboratorian. In addition, the statistical logic is not adequately explained for laboratory use, further limiting the usefulness of the guideline.

- **See the response to Comment 12.**

14. It is not necessary for laboratories to be able to verify the linearity of their methods with such statistical accuracy—visual evaluation of a graphical plot is often all that is necessary. However, because the document makes several statements about the statistical analysis being the only acceptable means of evaluating or establishing linearity, if this document is approved, laboratories all over the U.S. will be cited by their inspectors for failure to perform their linearity studies according this guideline. In general, it is unacceptable for a guideline to be prescriptive to users about the application of any protocol.

- **See the response to Comment 7.**

15. If the guideline is intended for manufacturers, it is too prescriptive for them as well. Manufacturers use various approaches to determine linearity, including an approach similar to that described in the guideline. In any case, whatever approach a manufacturer chooses to use, the average laboratory would never be expected to be able to repeat their studies, which are far more statistical than is necessary for the laboratory.

- **See the response to Comment 12.**

Title

16. The document emphasizes the use of a statistical method for the evaluation of linearity as the only "acceptable" approach. (This claim is not acceptable, as mentioned above.) The title needs to be changed to clarify the specific tool of this document. Also, this protocol is meant for the clinical laboratory, as documented by the section on the preferred samples to use. Change title to: *A Statistical Protocol for the Evaluation of the Linearity of Quantitative Analytical Methods for Use in Clinical Laboratories.* This title is consistent with the first line of the Purpose.

- **Please see the response to Comment 7.**

Foreword

17. 1st paragraph. Linearity does not have to be verified with a mathematical relationship. Our experience – and reports from our customers – is that visual evaluation is often all that is necessary. Delete "mathematical." The examples support the acceptable use of visual assessment.

- **See the response to Comment 12.**

18. 1st paragraph. Clinical linearity should also be considered: The assay is considered linear if the deviations are not clinically important.

- **See the responses to Comments 6 and 9.**

19. 2nd paragraph, 2nd sentence. Emphasize that dose-response relationship is not necessary for tests that are used for qualitative purposes, e.g., to determine the presence of a substance or characteristic. Otherwise, customers might ask for this relationship and they might expect a linear response curve. Edit sentence to: "…the dose-response curve to linear form if the assay is used for quantitative measurements."

- **The title has been modified to clearly indicate that the statistical protocol is for quantitative methods.**

20. Dose-response relationship terminology is usually specific to immunoassays, so it should be changed to be more general.

- **This "dose-response" phrase is used only in the Foreword, and is appropriately used as an example to reflect the general relationship between concentration (or activity) of the analyte (or measurand) and the measured response.**

21. 2nd paragraph. 3rd sentence. This is false. Interpolation can take place with nonlinear relationships. Delete the last sentence.

- **The working group agrees that it is possible to interpolate with nonlinear relationships; however, interpolations by clinicians are usually performed mentally, and are made with an assumption of a linear relationship. The sentence remains as written.**

22. 7th paragraph, 2nd sentence. Syntax error; change "will" to "can be."

- **The suggested change has been made.**

23. "This revision." This section is not appropriate in the document itself. Suggest this be moved as a response to previous comments.

- **Due to the 20-year development of this project, the working group believes this explanation is important.**

24. 9th paragraph, 2nd sentence. Mentioning there are unresolved issues, without pointing out what they are, does not serve to have input on the controversial points. Comment: "This was a serious limitation by the committee – why by the "committee?" Projects are subcommittee activities. This is a NCCLS process issue that needs to be addressed.

- **The "Invitation for Comment in the Consensus Process" is a standard statement included in all proposed-level documents. It is intended to solicit input on the current document. This project, however, was recommended (according to NCCLS policy) for cancellation due an expired timeline and lack of subcommittee activity. The Area Committee on Evaluation Protocols accepted the project as an area committee activity and a working group was assigned (also according to NCCLS policy). A draft document was prepared and circulated twice to the area committee and area committee advisors and observers for comment. Significant changes were made to the EP6 draft; hence its release as a second edition proposed-level guideline. All previous subcommittee members were invited to participate, and many suggestions were incorporated into the current guideline.**

25. I wish to question the NCCLS's process for updating documents. Please refer these comments to the appropriate person at NCCLS. I was an Observer on EP6 for several years. And thought that I had remained so after publication of the last version in the expectation that this would serve to keep me notified of any future changes. Now I belatedly see that a P2 has been prepared and I was unaware of it. What happened to my Observer status? I was never notified that I had lost it, and thus could not rely on NCCLS to keep me apprised of developments on this document.

- **Please see the response to Comment 24.**

Purpose (Section 1.1 [now Section 2.1])

26. Change the Title to match the Purpose.

- **The title of the document has been changed to *Evaluation of the Linearity of Quantitative Analytical Methods: A Statistical Approach; Approved Guideline.***

27. The point that users need to understand and determine their own quality needs (for total error, bias, etc.) is critical to appropriate use of this protocol. Begin a new paragraph with this sentence: "Users should have…" to emphasize the importance of this message. Also, reference NCCLS document EP21—*Total Analytical Error for Clinical Laboratory Methods.*

- **The suggested change has been made.**

28. The document states that it covers a broad range of situations. However, this protocol does not take into consideration whole blood testing devices, especially when the analyte is quickly metabolized, e.g., glucose.

- **This document does not claim to assess linearity for every measurement procedure. It is assumed the user will know when the procedure is appropriate.**

29. Users should focus on setting goals for total error across the reportable range. It is not necessary for the user to establish goals for each component of error.

- **See the responses to Comments 5 and 8.**

30. Clarify why the linearity of a quantitative analytic method should be determined. It is not clear if the linearity is a part of performance evaluation for a quantitative test or it is the only way to evaluate the performance. Note that a perfect straight line may not necessarily mean the quantitative test is accurate. For example, Y=1+2X (X = true concentration; Y = test result) says the test result is systematically away from the truth. It seems the linearity only does not make much sense but linearity with slope=1 and interception=0 does. No discussion on testing slope=1 and interception=0 was found.

- **This procedure is intended for assessing linearity independently, as far as possible, from other components of bias and random error. The suggested procedure would be for assessing error relative to a particular model for error, and with a particular scheme of X-axis solutions.**

Definitions (Section 1.3 [now Section 2.3])

31. The term "least squares regression" is odd to both statisticians and nonstatisticians. The "least square" is a mathematical algorithm which can be used to estimate parameters of any kind of regression model. Suggestion: delete the term from Section 1.3 and avoid using it throughout the entire document.

- **A variety of statistical "regression" procedures are possible. This document specifies a particular, widely available procedure.**

Overview of the Experimental Approach (Section 1.4 [now Section 2.4])

32. Include solutions that are not prepared by dilution, as is "allowed" in other parts of the document. Edit: "…requirement that the dilution or formulation be equidistant."

- **The end of the third paragraph has been edited to read: "… and there is no requirement that the assumed values are obtained by dilution."**

33. Change "…measurements from 5 to 9 concentrations…" to "… measurements from 5 to 9 samples with varying concentrations…"

- **The suggested change was made.**

34. Change "…requirement that the dilutions be…" to "…requirement that the sample concentrations be…"

- **The suggested change was made.**

Scope (Section 1.5 [now Section 1])

35. This protocol has very little to do with ensuring that a laboratory is operating competently. Also, mentioning compliance with regulation is both incorrect (e.g., CLIA has no such requirement) and beyond the scope of any guideline. Delete the last phrase from: "…so laboratories can operate…"

- **The suggested change has been made.**

36. The comment that this protocol does not consider precision is contrary to other messages where it is stated that good precision enhances the utility of the protocol. Clarify for consistency.

- **The first bullet has been edited to read: "This protocol is to assess linearity, isolated as much as possible from conditions of precision and trueness. It is understood that poor precision will hinder an effective assessment of linearity, therefore, a check for poor repeatability is included."**

37. Remove "so laboratories can operate competently and comply with regulations." There is no CLIA requirement to perform linearity testing.

- **See the response to Comment 35.**

38. Remove "…even though linearity is useful in assessing accuracy and the estimation of linearity is affected by imprecision." Although inaccuracy may be caused by nonlinearity, linearity by itself has nothing to do with accuracy.

- **The working group disagrees. The degree of nonlinearity is a continuum from negligible to unacceptably large, and therefore can be an important component of accuracy. However, the first bullet in Section 1 has been revised in response to Comment 36.**

39. Sample names should be used. Revise text to read: "These experiments should use samples (serum or plasma and urine) with matrix appropriate to the specimens being analyzed."

- **The parenthetical note "(serum, plasma, urine, etc.)" was added to the text.**

40. The Scope should specify specific domains of application where a linear range is expected; for example, to determine the concentration of analytes in clinical biochemistry, measuring the concentration of specific antibodies to a particular microorganism (rubeola, hepatitis B) in enzyme immunoassays.

- **The technique is general and can be applied to many assays for quantitative measurements.**

41. It is not clear for whom this procedure is intended. It says laboratories; however, it is too complicated for the moderately complex labs to comply with. If this is a procedure meant for the manufacturer, it should be stated as such.

- **The Abstract on p. i clarifies the intended audience for the document.**

42. Verifying linearity is not a U.S. requirement. If this is meant to provide a protocol to run reportable range studies, it should say that. Remove "and comply with regulations."

- **See the response to Comment 35.**

Duration of the Experiment (Section 2.2 [now Section 3.2])

43. Change "…in as close a time interval" to "in as short a time interval."

- **The suggested change has been made.**

44. As written, the text says "when" the experiment should be done, not how long it will take. Duration infers a clarification of "how long." At the end of the sentence, change the period to a comma and add the following: "..., and should take no more than one working day to complete."

- **The suggested change has been made, although the wording was changed to indicate that the restriction to a single day may not be practical for all analytes being tested.**

45. Remove "…after the device familiarization period has been completed." This statement is not necessary.

- **The working group believes this is an important caution.**

Evaluating Experimental Conditions (Section 3 [now Section 4])

46. Remove "…established by dilution." It is not required to prepare samples for linearity testing by dilution.

- **This is clarified in the first part of the sentence.**

Specimen Requirement (Section 3.1 [now Section 4.1])

47. Remove "including manufacturers." Manufacturers are not in the scope of this document.

- **The procedures are appropriate for anyone that develops or significantly modifies laboratory test and measurement methods. Please see the Abstract on p. i.**

48. Remove the word "considered." You must be definitive on the minimum number of points.

- **The suggested change has been made.**

49. The statement about the minimum number of points is equivocal. Delete "considered."

- **The suggested change has been made.**

Matrix Effects (Section 3.2 [now Section 4.2])

50. Add text: "All specimens should be free of interferences *and conformity of reactivity* that the method instructions identify (such as that which may be seen with icterus, hemolysis, or lipemia)..."

- **The working group disagrees; but the text has been edited to read "All specimens should be free of known interferences that the method instructions identify…."**

Hierarchy of Acceptable Matrices (Section 3.3 [now Section 4.3])

51. There is no mention of whole blood. Provide possible linearity, such as spiking whole blood with glucose solutions, or using appropriate commercial controls with five predefined glucose levels.

- **"Whole blood" has been added to the matrices mentioned within the parentheses. The other issues are adequately explained in the sections that follow.**

Pool Diluted with Treated Materials of Low Concentration or Treated Pool Material (Section 3.3.4, now 4.3.4)

52. Add "Caution: In case of preparation of a low concentration, the components of matrix do not change without measuring analytes (such as cannot be used for dilution with buffer solution to the serum."

- **The working group agrees. A fourth sentence has been added to Section 4.3.4: "It is important to retain a constant matrix rather than achieve low levels simply by dilution."**

Aqueous Solutions (Section 3.3.8 [now Section 4.3.8])

53. This is unclear. Edit (suggested): "Aqueous solutions are the least desirable for use with this protocol because they deviate the most from the matrix of the samples used in clinical laboratories."

- **The working group believes the current text has appropriate cautions.**

54. Requiring (or strongly recommending) the use of materials of the highest quality is too restrictive. Since the protocol can be used without knowing the absolute concentrations of measurand, purity is not a requirement as long as there are no interferences that can produce apparent non-linearity, i.e., false rejection of performance. Edit: "Although materials of high purity are preferable since they minimize the possibility of interferences that could affect interpretation of the results, materials of lesser purity could be acceptable."

- **The suggested change has been made.**

Selection of Materials Used to Supplement Samples (Section 3.4 [now Section 4.4])

55. No suggested considerations for appropriate means of producing a spiked sample of reasonable matrix. Although it may be obvious to try to identify a negative patient pool for the analyte, this is not always possible. Directly producing a sample at the upper limit of many linear ranges would require either weighing a minute amount of material or dissolving into a large volume of matrix. These concerns should be mentioned as requiring consideration when using this technique to make the high sample. Although protocols may be suggested or proffered, if the intent is not to make this document into a cookbook, an awareness of these considerations should be given.

- **It is not possible to describe all ways to make samples. However, a sentence about retaining a constant matrix was added to Section 4.3.4. (See Comment 52.)**

56. Table 1. Is sodium carbonate really the best material for $C0_2$, rather than bicarbonate? Should both or either be suggested, or is there some chemical issue with the use of bicarbonate?

- **The working group agrees. Sodium bicarbonate has been added to Table 1 for Carbon dioxide (CO_2).**

57. Table 1. Any time an NCCLS document makes recommendations such as this, they become *de facto* standards. This table is prescriptive, and not even very enlightening. Mostly it lists "purified enzyme," or "purified (analyte)," with few other specific recommendations. Delete Table 1. Section 3.4 is sufficient without it, or could be made so with another sentence or two on material selection criteria.

- **The working group disagrees. Table 1 is a useful guide and the second paragraph of Section 4.4 provides guidance.**

Sample Preparation and Value Assignment (Section 3.7 [now Section 4.7])

58. Last paragraph. A caution should be added. If a piston-stroke pipette is used against serum or plasma sample, it will obtain an inadequate sample volume. It is better to make the mixture by dilution with gravimetry. Please add the following: "Caution: Dilutions are more accurate if prepared using gravimetry instead of using a piston-stroke pipette."

- **The following has been added to the second paragraph of Section 4.7: "This example uses a dilution-based scheme. In some laboratories, a gravimetric scheme may be preferred. Regardless of the procedure for preparing solutions, the preparer should use due caution."**

59. Although the protocol(s) for preparing equally spaced samples are straightforward, they require the use of multiple fixed-volume pipettes or measuring pipettes which makes the analyte value for most points dependent on the precision and accuracy of two pipettes per sample. For the 5-sample pattern, I use equal-volume additions in this sequence:

    #1: Low sample (or pool)
    #2: 1 part Low + 1 part Midpoint
    #3: Midpoint (1 part Low + 1 part High)

#4 1 part High + 1 part Midpoint
#5: High sample (or pool or spiked sample)
(The midpoint mixture must be made first, followed by #2 or #4 in arbitrary order). This allows the same fixed-volume pipette to be used to make each level, taking pipette accuracy out of the picture and just making the error dependent on pipette precision. A similar process can be done for 9-sample patterns, but keeping it straight is mind-boggling. This cannot be done for other patterns being recommended in the guideline.

- **The working group agrees that the suggested procedure would have the stated effect. However, there are several ways to make solutions to use in this protocol. Rather than try to list them all, the working group elected to recommend a single, unified approach for many different levels.**

Data Collection (Section 3.8 [now Section 4.8])

60. Section 3.8 describes the appropriate number of replicates for each evaluation purpose; these replicate numbers are insufficient for some test methods. Therefore, it would be helpful to have guidance within the document to help the laboratory select the appropriate number of replicates for the test method under evaluation. This would be based upon the known variance of the method at the sample concentrations under consideration.

    For example, replicate %CVs of <5% are common in immunoassays, and so 2-3 replicates per dilution level is appropriate in assessing linearity of these assays. However, in the increasingly common nucleic-acid amplified assays, such as PCR-based assays, the replicate %CVs are much higher, usually >20% and as high as 45%. For these assays, 2 replicates per dilution level is insufficient to evaluate assay linearity.

- **The experience of the working group is that the suggested number of replicates is appropriate for most methods, and minimizes the number of tests. However there are situations where more replicates are needed to provide reliable mean values. The following paragraph was added to Section 4.8:** "**The number of replicates should be sufficient to produce a reliable estimate of the concentration at each level. For some analytes, at some concentrations, that may require 3-5 replicates. Users should use their best judgment about the number of replicates needed. The procedures in this protocol are not adversely affected by the use of different numbers of replicates at different levels."**

61. Bullets 1 and 2. Not clear why establishing the linear range and validating claims for "in-house" methods are different. Both are establishing claims.

- **There are a variety of purposes for which the procedure might be used, with varying needs for in-house assessment of modified methods. The stated options are intended to suggest different levels of action, depending on particular concerns.**

62. The statement is prescriptive, without any justification. There is no accepted standard for the number of samples required for each of these objectives. Edit: "The following are the number of samples suggested for different purposes."

- **The suggested change has been made.**

Preliminary Data Examination (Section 4.1 [now Section 5.1])

63. Add instruction for plotting the data for visual inspection after step (1). Add specific instruction, including specifics, e.g., if the independent variable is made from admixtures or from known or assigned concentrations, whether all results (dependent variable) are plotted or the mean of replicates, etc.

- **Additional text was added to Section 5.1 (2) to further clarify the plotting procedure.**

64. Visual inspection of the plot must be acceptable for confirming linearity, if so desired, just as it is used to reject linearity (meaning, do not bother to perform the complex statistical analysis). Change the latitude of the protocol *or* change the title and scope for this to be a *statistical* protocol for the evaluation of linearity. The examples support the acceptable use of visual assessment.

- **A visual assessment is useful to detect important nonlinearity, but a visual assessment may not be sufficient to confirm adequate linearity. The title of the document has been changed to deal with this concern.**

Outlier Test (Section 4.2 [now Section 5.2])

65. Outliers can also come from assumed mistakes, e.g., intermittent air bubble.

- **The text has been edited to read: "Outliers are test results that reflect mistakes (clerical error, system glitches, etc.) whether proven or reasonably assumed."**

66. Add ", if possible" to the end of the paragraph.

- **The suggested change has been made.**

Polynomial Regression (Section 4.3.2 [now Section 5.3.2])

67. This section needs substantial work to make it comprehensible for the well-trained laboratory professional. There are many questions that need to be addressed, including definitions and what the analysis represents in meaningful words. It is *not* acceptable practice to refer users to software programs if they are not instructed about what each parameter is supposed to mean. This is especially critical for establishing effective acceptability criteria to compare against.

- **Additional text was added to this section to better explain the statistical analysis.**

68. This is the technical and statistical guts of the entire document. And it has been given very short shrift. More text was present on Purpose in Section 1 than here. Needs to be expanded and made clearer for laboratory personnel who will be familiar with statistics, but not necessarily expert in these particular ones. Rewrite section for greater detail, instruction, and clarity for the nonexpert user.

- **Please see the response to Comment 67.**

69. "Obtain the Standard Error of the slope…and apply the t-test" is insufficient. This is a nontrivial exercise and needs greater detail, instruction, and clarity. Rewrite section for greater detail, instruction, and clarity for the nonexpert user.

- **Please see the response to Comment 67.**

70. Explain "standard error of the slope for each nonlinear coefficient." What if the "t-test" fails? What does it mean if b_2 passes, but b_3 fails? Or vice versa? Or any other combinations? Add these terms to Definitions. Explain the importance of the parameter and how they can be used with the acceptance criteria that is established by the laboratory.

- **Please see the response to Comment 67.**

71. Add definitions of the statistical terms, keeping the specific context of this protocol in mind, e.g., "Rdf is the number of degrees of freedom consumed by…" How did *Rdf* become = 4 in the example? Define and *illustrate*: *df; Rdf*.

- **Please see the response to Comment 67.**

72. First paragraph. Change "…the intervening solutions would" to "…the intervening solutions could"

- **The text has been changed from "…the intervening solutions would…" to "… equidistant solutions would…"**

73. Section 4.3.2, page 11 and Appendix B, page 18. The text (page 11) indicates a one-sided t-table was attached (Appendix B). But the Appendix B t-table apparently presented the two-sided t values.

- **The text was changed to state that a two-sided test is used.**

Degree of Nonlinearity (Section 4.3.3 [now Section 5.3.3])

74. There is no instruction on how to "pick" the polynomial. Does this mean to select the equation where b_2 is the smallest? Rewrite for clarity for the well-trained laboratory professional.

- **The selection is based on the model with the lowest standard error of regression.**

75. The term $p(x_i)$ is not explained. Is this the measured result? If so, what should be done with the duplicate measurements? Take the average? Rewrite for clarity for the well-trained laboratory professional.

- **This term is described as "the value of the best-fitting polynomial at point x_i." and point x_i is the known or assigned value for that sample.**

76. How would this evaluation be conducted if the laboratory uses admixtures and does not know or does not estimate the actual concentrations of measurand? Rewrite for clarity for the well-trained laboratory professional. Also, the issue of appropriate limits is not trivial. Are these parameters to be compared to bias or total error? Even if bias is used, does that then imply that linearity is a fair judge of bias?

- **The document states that if the actual values are unknown, the relative concentrations are sufficient. The discussion on appropriate limits was extensively revised.**

77. DL_i needs to be explained. Is the deviation to be evaluated only at the points that solutions were tested? Or should points be interpolated along the polynomial? If so, how many points are needed? Rewrite for clarity for the well-trained laboratory professional. This is a good case for why visual inspection is a very effective tool for evaluating linearity. Also, the examples support the acceptable use of visual assessment. (See comment for Appendix C)

- **The text has been revised to better explain DL_i. Please see the response to Comment 64 regarding visual inspection.**

78. This protocol is specific for systems, not just for instruments (this probably comes for the fact that the CAP Linearity Survey comes from the Instrumentation Resource Committee). Edit: "…contact the service representative for your system for help…"

- **The second sentence in Section 5.3.3(1) has been edited to read: "You may wish to contact the representative for your system or instrument for help with this investigation."**

79. First paragraph. "This statistic is the average difference between…" is not true. It is not a simple average. Suggest: "This statistic is a measure of the difference…."

- **The suggested change has been made.**

80. Sy•x needs more information on how to obtain or derive this value. Rewrite section for greater detail, instruction, and clarity for the non-expert user.

- **The text was revised to better explain the statistical analysis, and the user can consult the basic statistical texts cited such as those listed in References 7-9.**

81. Again, needs more clarity on how to calculate DL_i. Explain what is a $p(x_i)$. Clarity of Section 4.4 is much better than 4.2 or 4.3 and could be used as a better, though not great, model of how to describe these calculations.

- **The text was revised to better explain the statistical procedures. See the responses to Comments 75 and 77.**

82. Second paragraph. "This is a measure of the difference between the nonlinear model and the best-fit straight line,…" Suggest changing to: "This is a measure of the difference between the best-fit nonlinear model and the linear model,…" Only one straight line is fitted and several nonlinear models were fitted per Section 4.3.2.

- **The suggested change has been made.**

Considerations for Random Error (Section 4.4 [now Section 5.4])

83. Revision for the entire section is needed. The large part of this section is about calculating random error. In the first paragraph, it's said that the random error can be calculated easily by hand, but the formula given in the same section cannot be used to do so for the data with 3 replicates though the calculation for both 3 and 4 replicate data are fairly easy. Suggestion: replace the SD formula for 2 replicate data on page 13 with the generic within-level SD formula. In addition, I don't understand how to calculate the SD using the CV as suggested on page 13.

- **For more than two replicates, the protocol suggests using ANOVA results or other software procedures. Since the procedure for two replicates is a simple adaptation, it is provided.**

  The sentence following the formula has been changed to: "r_{i1} and r_{i2} can be either the actual results of the procedure, or they can be expressed as percentages of the solution mean (although the same units must be used for all levels). If percentage differences are used, then the result is a CV_r, not SD_r."

84. "r_{i1} and r_{i2} can be either actual results, or expressed as percentages of the level mean (CVr)" should also note that actual results and percentages cannot be commingled in the same calculation. Either one or the other must be consistently used in calculation of CV_r. Paraphrase comment.

- **The text has been edited to read, "…although the same units must be used for all levels."**

85. The statement "pooled differences between L" is neither defined nor explained. Edit by including in the Definitions, and provide clarity for the well-trained laboratory professional.

- **The text has been modified to better explain the estimate of pooled differences.**

86. Although the example is for duplicates, and more replicates are "permitted," there is inadequate discussion about what to do for all the inputs that are required; e.g., degrees of freedom. Modify to either explain other test designs or confine the protocol to duplicates. Note especially, the equation for SD_r, where "only" duplicate (is that what $_{i1}$ and $_{i2}$ mean?) measurement is the example.

- **A formula was added to describe the calculations with more than two replicates.**

87. "Repeatability" is defined in international standards, so it is inappropriate to say, "often called." I don't see where it adds anything to the discussion. Edit this sentence to remove the term "repeatability."

- **Terminology has been modified consistent with the NCCLS policy for an evolutionary change toward harmonization with international practices. Please see the "Note on Terminology" in the Foreword.**

88. These sentences contain subjective instructions with no help in determining what is "reasonably equal" or is "much higher." Edit with more clarification. This document needs to be understood by the first-time user. Everyone who uses it will be a "first-timer," so it won't be possible for those people to make a judgment based on their experience.

- **Procedures for assessing the size of the error are given elsewhere. The determination of suitability must be subjective, based on the needs of the user.**

89. Directing users to "other spreadsheets" for analysis of variance is not helpful for a standard protocol. Provide better instructions (and explanations) or discard the use of more replicates.

- **See the response to Comment 86.**

90. Fourth paragraph. Clarify the term "the average could be compared…" Edit: "the average of the replicates for each of the solution concentrations tested…"

- **The term "average" referred to the average difference between replicates. The sentence has been edited to read: "Compare SD_r with the goal for repeatability either as is, or expressed as a percentage of any concentration."**

91. Why is SD_r compared to the goal, and not twice or three times SD_r? Explain and clarify the relationship of SD_r to the goal, either total error or precision, and what it could mean.

- **The goal for precision should be compared with SD_r. See the responses to Comments 98, 99, and 100.**

92. When variability is not constant or proportional, the user will have a difficult time making an assessment of linearity, and pointing out more limitations and referring to still additional software is not helpful (especially without guidance on how or where it can be obtained). Delete or modify so the information is helpful.

- **This document assumes understanding of basic concepts of linear relationships; further explanation is unnecessary.**

Graphical Approaches (Section 5.1 {now Section 6.1])

93. The statements about acceptance and rejection are beyond the authority of a guideline. An evaluation protocol *cannot* declare that visual inspection is insufficient for acceptance. Delete from ", but not acceptance,…or check claims." The fact is that many manufacturers (my company included) *often use visual evaluation for acceptance!* And with success!

- **Visual assessment by experts who are familiar with possible sources of important errors, and experienced in spotting such errors relative to the requirements of the instrument may successfully confirm that a system is in control. However, an objective assessment of linearity relative to goals for error requires an objective statistical analysis.**

94. Visual plots alone can be sufficient to check claims. If no clinically relevant deviations are noted within the reportable range, then the linearity is sufficient. This can easily be established by visual evaluation of the plot.

- **See the response to Comment 93.**

Verification of Claims for Linear Range (Section 5.2 [now Section 6.2])

95. This paragraph, and the concept behind it, is disingenuous. NCCLS documents become *de facto* standards for regulatory and other purposes. To say that many methods may be used to establish linearity, including this one, is technically true,

but begs the question that hardly anyone will now use other methods once this is published. Alternatively, list some example alternative methods that would be considered equally valid. Our challenge is to create a consensus document that everyone can use and live with. 1) Make EP6 as good as we can with true consensus that everyone can use and live with. 2) List example alternative linearity tests that would be considered equally valid.

- **Alternatives are presented in the introductory comments and options are discussed throughout the document. Also, please see the responses to Comments 2 and 3.**

96. "Operation" usually means many or all aspects of a process. This protocol evaluates the analytical system. Edit "operations" to "system."

- **The suggested change has been made.**

97. There is no justification to support the statement that the "linear claims should be based on a statistical procedure…" The fact is that many manufacturers (my company included) *often use visual evaluation for acceptance!* And with success!

- **The sentence has been deleted. Please see the response to Comment 93.**

Allowable Error and Goal Setting Section 5.3 [now Section 6.3])

98. The document instructs the laboratory to establish goals for "allowable error due to nonlinearity." It also recommends that users "have an understanding of their needs for total error, bias, random error (or imprecision), and nonspecificity (or interferences)." Although it is an interesting endeavor to establish goals for each of these components, it is not necessary. It really does not matter if error is due more to precision or bias or nonlinearity. It only matters that the total error is within acceptable limits.

- **This document covers one important source of uncertainty. Proper understanding and reduction of uncertainty requires breaking it into components. See the responses to Comments 99 and 100.**

99. None of these references specifically address establishing goals for linearity. Further explanation is needed if they are included. Establishing goals for components of error (starting from Allowable Error) is not a trivial exercise. Should linearity goals be based on or be equivalent to bias goals? (The Ehrmeyer and Laessig article could be helpful here.) What parameters are to be used for goal comparison, b_2 or b_3 or something else?

- **As the document states, the references provide assistance with setting goals for the various components of accuracy, with different models. The working group agrees that goal setting is complex, and the following bullets were added at the end of Section 6.3 to further define considerations for setting goals for measurement error, bias, and nonlinearity:**

  - **Goals for linearity should be derived from goals for bias, and should be less than or equal to those goals.**
  - **Goals for bias should be less than or equal to goals for measurement error.**
  - **When the concentrations of the test samples are unknown, the goals for linearity should be in relative (percentage) units.**

  The stated references provide guidance on establishing goals for error and for various components of that error. Although goals for nonlinear error are not specifically addressed, the information in the references along with guidelines in this document will allow such determinations. The issue of setting goals for various components is a subject for a future document.

100. Remove "This should include goals for the various sources of error, such as precision, bias, and nonlinearity." Labs do not need to set separate goals for these error components. Only total error goals are needed. If unacceptable total error occurs, then they can troubleshoot the system by looking at the components.

- **The first sentence was edited to read "The laboratory must determine its own goals for total acceptable error for every analyte." In the third sentence, "This should include goals for the various sources…" was replaced with "Requirements for measurement error can be used to determine goals for the various sources…"Also, see the response to Comment 99.**

101. May I suggest "The laboratory must determine its own goals for error [for each analyte]. The goals should be based on the [biovariability of the analyte in normal individuals], the needs of the clients of that laboratory and an understanding of ... (my additions in square brackets).

- **The current explanation is sufficiently clear. The bioavailability of the analyte in normal individuals is unnecessary.**

Suggested Wording of Claims (Section 5.6.1 [now Section 6.6.1])

102. Wording of claims is beyond the scope of evaluation documents. Linearity is a means of verifying the reportable range (which is a required element of a manufacturer's claim in many regions of the world), so linearity might not be mentioned at all. In addition, there is already a proposed NCCLS document dealing with Claims. Delete.

- **The text states that alternative wording is possible.**

103. Verbiage for claims should NOT be dealt with in this document. This is entirely too prescriptive. A number of other (often conflicting) documents and regulatory agencies already deal with claims verbiage. Don't make the situation worse. It's unnecessary. Besides, who defines "acceptable" as in the phrase "… developer provides acceptable alternative wording?"

- **The word "acceptable" was removed and the following was added to the end of the sentence: "alternative wording that includes the range of linearity and maximum difference."**

104. Manufacturers are not likely to report the goal or measured max difference. This number is not relevant. Only the total error is relevant, as established by method comparison.

- **See the response to Comment 102.**

105. Remove section. Manufacturers do not and cannot prescribe clinically allowable differences in their labeling. It is up to the manufacturer to provide truthful information regarding the test systems' performance. Then it is up to the user to determine if that performance meets their needs.

- **See the response to Comment 102 and readers should refer to the most current version of NCCLS document EP11—*Uniformity of Description of Claims for* In Vitro *Diagnostic Tests*. The suggested wording is only relevant if a claim is made that linearity follows this protocol; there is no need to claim clinically allowable differences — linearity can be claimed to the largest nonlinear difference, with no statement of clinical appropriateness.**

106. Laboratory professionals who refer to the available literature (much of it referenced here) understand that goals might change according to concentration and the use of the measurand. This reminder should be deleted. It does not provide help with the difficult task of establishing goals (see comments above).

- **A reviewer suggested this option for the user, and the working group agreed it was a useful suggestion. Text was added to the end of Section 6.3 to state that relative goals (percentage) should be used in situations where the concentrations are not known.**

Dilution Schemes (Appendix A)

107. A gravimetrical procedure should be included. Add the following:

Dilution scheme using gravimetry:

S=5 samples
1:Low(L)1.0300g
2:0.7725L+0.2575H
3:0.5150L+0.5150H
4:0.2575L+0.7725H
5:High(H)1.0300g

- **The working group disagrees. When preparing equally spaced solutions, the gravimetric procedure would not assure equal spacing. However the gravimetric procedure can be used to prepare solutions with targeted concentrations rather than assumed equal spacing.**

108. Why do you enumerate all the possibility from S=5 to S=11 samples? If you want to save space in the article, you could use a generalization of the procedure that enables the people to easily go to a larger sample and only put two examples. Of course, if space is not an issue, please disregard this comment.

- **The working group believes the current approach is adequate.**

109. We could read the first paragraph already present on page 17 and add another paragraph saying:

The concentration of the low (L) and high (H) sample can be calculated with this formula:

Let's say Increment (I) = 1/(S-1)

For the low concentration (L) start at 1.00 and subtract I for each level more than L.

For the high concentration (H) start at 0 and add I for each level more than L.
Here are two examples with S=5 and S=11 samples. You can calculate the others.

S=5 samples	Calculated with:
I = 1/(S-1) = 1/(5-1) = 0.25	
1: Low (L)	
2: 0.75L + 0.25H	(1.00 – 1* I)L + (0.00 + 1* I)H
3: 0.50L + 0.50H	(1.00 – 2* I)L + (0.00 + 2* I)H
4: 0.25L + 0.75H	(1.00 – 3* I)L + (0.00 + 3* I)H
5: High (H)	

S=11 samples	calculated with:
I = 1/(S-1) = 1/(11-1) = 0.10	
1: Low (L)	
2: 0.90L + 0.10H	(1.00 – 1* I)L + (0.00 + 1* I)H
3: 0.80L + 0.20H	(1.00 – 2* I)L + (0.00 + 2* I)H
4: 0.70L + 0.30H	(1.00 – 3* I)L + (0.00 + 3* I)H
5: 0.60L + 0.40H	(1.00 – 4* I)L + (0.00 + 4* I)H
6: 0.50L + 0.50H	(1.00 – 5* I)L + (0.00 + 5* I)H
7: 0.40L + 0.60H	(1.00 – 6* I)L + (0.00 + 6* I)H
8: 0.30L + 0.70H	(1.00 – 7* I)L + (0.00 + 7* I)H
9: 0.20L + 0.80H	(1.00 – 8* I)L + (0.00 + 8* I)H
10: 0.10L + 0.90H	(1.00 – 9* I)L + (0.00 + 9* I)H
11: High (H)	

- **This procedure is unnecessary. With equally spaced samples, there is no need for estimate concentrations. Users can use the observed concentrations.**

Examples (Appendix C)

110. In Appendix C, Figures 1 to 6 should use different line style in their graphics so we can read them more easily. There's no point in using a legend when «1st» and «2nd» use the same symbols.

- **Figure C1 was edited for clarification. Figures C2-C6 are sufficiently clear.**

111. Even with the explanations and analysis of the data, it is not clear to me that this complex analysis adds anything more that the visual interpretation with the goals shown. More supportive instruction is needed. This adds to the argument that either visual or statistical analysis might serve the purpose.

- **Text has been added to better explain the examples. The two examples are intended to demonstrate the protocol on (1) a dataset where there is obvious and significant nonlinearity; and (2) a dataset with nonlinearity that exceeds criteria on the full range, but meets the laboratory's criteria on a reduced range. These examples demonstrate the ability to assess nonlinearity and repeatability with objective quantitative measures that can be either compared with goals, or used to better understand this important component of measurement error.**

112. Example 1. In the first table, the mean and slope of the 4th dilution should be 340 and 9.0, respectively. The slope of the 5^{th} dilution should be 66.5."

- **The working group appreciates the comment. The error has been corrected.**

113. The method described is cumbersome and requires repeating calculations many times. This repeated analysis can be eliminated by establishing the regression line to have zero bias at the mean normal result, and then looking to see where the maximum allowable bias occurs in both directions from the mean normal (lower and upper limits).

- **The suggested procedure would be useful for estimating method bias at different levels. It would be preferred for testing measurement error and for investigating a particular hypothesis for nonlinearity; but it is beyond the scope of this document, which is to measure linearity separate from bias.**

Summary of Delegate Comments and Working Group Responses

EP6-A: *Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline*

General

1. These statistical based approaches to trace blood Pb data reduction and interpretation are very useful for their intended purpose. I suggest that such algorithms be used for most newly created calibration curves when questions arise about whether or not a given calibration curve is within the so-called linear dynamic range for any atomic absorption spectrophotometer and possibly for any of our gas chromatographs that measure trace organics in our laboratory. I doubt, however, that this much computation is necessary in routine analytical work. To understand this guideline requires knowledge of statistics with applicability to analytical data. I am not convinced that many public health lab workers would be equipped to understand as well as to apply these mathematical concepts without formalized training.

- **The working group appreciates the comment. Also, please refer to the responses to Comments 2 and 3 in the Summary of Comments and Working Group Responses.**

2. The procedure described in this document is too difficult and cumbersome for routine lab use.

- **Please see the responses to Comments 2 and 3 in the Summary of Comments and Working Group Responses.**

3. Statistics need to be explained more fully. Stating that perimeters are "available from the regression program output" is not sufficient. Descriptions should be sufficient enough so that users can set up formulas in spreadsheets.

- **Please see the responses to Comments 2 and 3 in the Summary of Comments and Working Group Responses.**

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 through a gap analysis. The approach is based on the model presented in the most current edition of NCCLS 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. 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

EP6-A 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 next page.

Documents & Records	Organization	Personnel	Equipment	Purchasing & Inventory	Process Control	Information Management	Occurrence Management	Assessment	Process Improvement	Service & Satisfaction	Facilities & Safety
					EP9 EP10 EP14 EP15 EP21				EP7		M29

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

Related NCCLS Publications*

EP5-A **Evaluation of Precision Performance of Clinical Chemistry Devices; Approved Guideline (1999).** This document provides guidance for designing an experiment to evaluate the precision performance of clinical chemistry devices; recommendations for comparing the resulting precision estimates with manufacturers' precision performance claims and determining when such comparisons are valid; as well as manufacturer's guidelines for establishing claims.

EP7-A **Interference Testing in Clinical Chemistry; Approved Guideline (2002).** This document provides background information, guidance and experimental procedures for investigating, identifying, and characterizing the effects of interfering substances on Clinical Chemistry test results.

EP9-A2 **Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline – Second Edition (2002).** This document addresses procedures for determining the bias between two clinical methods or devices, and for the design of a method comparison experiment using split patient samples and data analysis.

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

EP14-A **Evaluation of Matrix Effects; Approved Guideline (2001).** This document provides guidance for evaluating the error or bias in analyte measurements that is due to the sample matrix (physiological or artificial) when two analytical methods are compared.

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

EP21-A **Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline (2003).** This document provides manufacturers and end users with a means to estimate total analytical error for an assay. A data collection protocol and an analysis method which can be used to judge the clinical acceptability of new methods using patient specimens are included. These tools can also monitor an assay's total analytical error by using quality control samples.

M29-A2 **Protection of Laboratory Workers from Occupationally Acquired Infections; Approved Guideline—Second Edition (2001).** Based on U.S. regulation, this document provides guidance on the risk of transmission of hepatitis viruses and human immunodeficiency viruses in any laboratory setting; specific precautions for 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 April 2003)

Sustaining Members

Abbott Laboratories
American Association for Clinical Chemistry
Beckman Coulter, Inc.
BD and Company
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 of Hematology
American Society for Microbiology
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.
Ontario Medical Association Quality Management Program-Laboratory Service
RCPA Quality Assurance Programs PTY Limited
Sociedade Brasileira de Analises Clinicas
Sociedad Espanola de Bioquimica Clinica y Patologia Molecular
Taiwanese Committee for Clinical Laboratory Standards (TCCLS)
Turkish Society of Microbiology

Government Members

Association of Public Health Laboratories
Armed Forces Institute of Pathology
BC Centre for Disease Control
Centers for Disease Control and Prevention
Centers for Medicare & Medicaid Services/CLIA Program
Centers for Medicare & Medicaid Services
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
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 Laboratories, MediSense Products
Acrometrix Corporation
Alifax S.P.A.
Ammirati Regulatory Consulting
Anaerobe Systems
A/S ROSCO
Asséssor
AstraZeneca Pharmaceuticals
Avant Immunotherapeutics, Inc.
Aventis
Axis-Shield POC AS
Bayer Corporation – Elkhart, IN
Bayer Corporation – Tarrytown, NY
Bayer Corporation – West Haven, CT
Bayer Medical Ltd.
BD
BD Biosciences – San Jose, CA
BD Consumer Products
BD Diagnostic Systems
BD Italia S.P.A.
BD VACUTAINER Systems
Beckman Coulter, Inc.
Beckman Coulter, Inc. Primary Care Diagnostics
Beckman Coulter K.K. (Japan)
Bio-Development SRL
Bio-Inova Life Sciences International
Bio-Inova Life Sciences North America
BioMérieux (NC)
bioMérieux, Inc. (MO)
Biometrology Consultants
Bio-Rad Laboratories, Inc.
Bio-Rad Laboratories, Inc. - France
Blaine Healthcare Associates, Inc.
Bristol-Myers Squibb Company
Canadian External Quality Assessment Laboratory
Chiron Corporation
ChromaVision Medical Systems, Inc.
Clinical Design Group Inc.
Cognigen
Community Medical Center (NJ)
Copan Diagnostics Inc.
Cosmetic Ingredient Review
Cubist Pharmaceuticals
Dade Behring Inc. - Deerfield, IL
Dade Behring Inc. - Glasgow, DE
Dade Behring Inc. - Marburg, Germany
Dade Behring Inc. - Sacramento, CA
Dade Behring Inc. - San Jose, CA
David G. Rhoads Associates, Inc.
Diagnostics Consultancy
Diagnostic Products Corporation
Eiken Chemical Company, Ltd.
Elan Pharmaceuticals
Electa Lab s.r.l.
Enterprise Analysis Corporation
Essential Therapeutics, Inc.
EXPERTech Associates, Inc.
F. Hoffman-La Roche AG
Fort Dodge Animal Health
General Hospital Vienna (Austria)
Gen-Probe
GlaxoSmithKline
Greiner Bio-One Inc.
IGEN Inc.
Immunicon Corporation
Instrumentation Laboratory
International Technidyne Corporation
IntraBiotics Pharmaceuticals, Inc.
I-STAT Corporation
Johnson and Johnson Pharmaceutical Research and Development, L.L.C.
LAB-Interlink, Inc.
Laboratory Specialists, Inc.
Labtest Diagnostica S.A.
LifeScan, Inc. (a Johnson & Johnson Company)
Lilly Research Laboratories
Medical Device Consultants, Inc.
Merck & Company, Inc.
Minigrip/Zip-Pak
mvi Sciences (MA)
Nichols Institute Diagnostics (Div. of Quest Diagnostics, Inc.)
NimbleGen Systems, Inc.
Nissui Pharmaceutical Co., Ltd.
Nippon Becton Dickinson Co., Ltd.
Norfolk Associates, Inc.
Novartis Pharmaceuticals Corporation
Ortho-Clinical Diagnostics, Inc. (Rochester, NY)
Ortho-McNeil Pharmaceutical
Oxoid Inc.
Paratek Pharmaceuticals
Pfizer Inc
Pharmacia Corporation
Philips Medical Systems
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.
SARL Laboratoire Carron (France)
Schering Corporation
Schleicher & Schuell, Inc.
Second Opinion
Seraphim Life Sciences Consulting LLC
Streck Laboratories, Inc.
Synermed Diagnostic Corp.
Sysmex Corporation (Japan)
Sysmex Corporation (Long Grove, IL)
The Clinical Microbiology Institute
The Toledo Hospital (OH)
Theravance Inc.
Transasia Engineers
Trek Diagnostic Systems, Inc.
Tyco Kendall Healthcare
Versicor, Inc.
Vetoquinol S.A.
Vysis, Inc.
Wyeth Research
Xyletech Systems, Inc.
YD Consultant
YD Diagnostics (Seoul, Korea)

Trade Associations

AdvaMed
Japan Association Clinical Reagents Ind. (Tokyo, Japan)
Medical Industry Association of Australia

Associate Active Members

31st Medical Group/SGSL (APO, AE)
67th CSH Wuerzburg, GE (NY)
121st General Hospital (CA)
Academisch Ziekenhuis-VUB (Belgium)
Acadiana Medical Laboratories, LTD (LA)
Advocate Healthcare Lutheran General (IL)
Akershus Central Hospital and AFA (Norway)
Albemarle Hospital (NC)
Allegheny General Hospital (PA)
Allina Health System (MN)
Alton Ochsner Medical Foundation (LA)
Anne Arundel Medical Center (MD)
Antwerp University Hospital (Belgium)
Arkansas Department of Health
ARUP at University Hospital (UT)
Armed Forces Research Institute of Medical Science (APO, AP)
Associated Regional & University Pathologists (UT)
Atlantic Health System (NJ)
Aurora Consolidated Laboratories (WI)
AZ Sint-Jan (Belgium)
Azienda Ospedale Di Lecco (Italy)
Baxter Regional Medical Center (AR)
Bay Medical Center (MI)
Baystate Medical Center (MA)
Bbaguas Duzen Laboratories (Turkey)
BC Biomedical Laboratories (Surrey, BC, Canada)
Bermuda Hospitals Board
Bo Ali Hospital (Iran)
Brooks Air Force Base (TX)
Broward General Medical Center (FL)
Cadham Provincial Laboratory
Calgary Laboratory Services
Cape Breton Healthcare Complex (Nova Scotia, Canada)
Carilion Consolidated Laboratory (VA)
Cathay General Hospital (Taiwan)
Cavan General Hospital (Ireland)
Central Peninsula General Hospital (AK)
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)
The Charlotte Hungerford Hospital (CT)
Children's Hospital (LA)
Children's Hospital (NE)
Children's Hospital & Clinics (MN)
Children's Hospital Medical Center (Akron, OH)
Children's Hospital of Philadelphia (PA)
Children's Medical Center of Dallas (TX)
CHR St. Joseph Warquignies (Belgium)
Clarian Health–Methodist Hospital (IN)
CLSI Laboratories (PA)
Commonwealth of Kentucky
Community Hospital of Lancaster (PA)
CompuNet Clinical Laboratories (OH)
Cook County Hospital (IL)
Cook Children's Medical Center (TX)
Covance Central Laboratory Services (IN)
Danish Veterinary Laboratory (Denmark)
Danville Regional Medical Center (VA)
Dean Medical Center (WI)
Department of Health & Community Services (New Brunswick, Canada)
DesPeres Hospital (MO)
Detroit Health Department (MI)
Diagnosticos da América S/A (Brazil)
Dianon Systems, Inc.
Dr. Everett Chalmers Hospital (New Brunswick, Canada)
Doctors Hospital (Bahamas)
Duke University Medical Center (NC)
Dwight David Eisenhower Army Med. Ctr. (GA)
E.A. Conway Medical Center (LA)
East Side Clinical Laboratory (RI)
Eastern Health (Vic., Australia)
Elyria Memorial Hospital (OH)
Emory University Hospital (GA)
Fairview-University Medical Center (MN)
Federal Medical Center (MN)
Florida Hospital East Orlando
Focus Technologies (CA)
Focus Technologies (VA)
Foothills Hospital (Calgary, AB, Canada)
Fresenius Medical Care/Spectra East (NJ)
Fresno Community Hospital and Medical Center
Frye Regional Medical Center (NC)
Gambro BCT (CO)
Gamma Dynacare Medical Laboratories (ON, Canada)
Geisinger Medical Center (PA)
Grady Memorial Hospital (GA)
Guthrie Clinic Laboratories (PA)

Hahnemann University Hospital (PA)
Harris Methodist Fort Worth (TX)
Hartford Hospital (CT)
Health Network Lab (PA)
Health Partners Laboratories (VA)
Highlands Regional Medical Center (FL)
Hoag Memorial Hospital Presbyterian (CA)
Holy Cross Hospital (MD)
Holmes Regional Medical Center (FL)
Holzer Medical Center (OH)
Hôpital du Sacré – Coeur de Montreal (Montreal, Quebec, Canada)
Hôpital Maisonneuve – Rosemont (Montreal, Canada)
Hôpital Saint-LUC (Montreal, Quebec, Canada)
Hospital for Sick Children (Toronto, ON, Canada)
Hospital Sousa Martins (Portugal)
Hotel Dieu Hospital (Windsor, ON, Canada)
Huddinge University Hospital (Sweden)
Hurley Medical Center (MI)
Indiana University
Innova Fairfax Hospital (VA)
Institute of Medical and Veterinary Science (Australia)
International Health Management Associates, Inc. (IL)
Jackson Memorial Hospital (FL)
Jersey Shore Medical Center (NJ)
John F. Kennedy Medical Center (NJ)
John Peter Smith Hospital (TX)
Kadlec Medical Center (WA)
Kaiser Permanente Medical Care (CA)
Kaiser Permanente (MD)
Kantonsspital (Switzerland)
Kenora-Rainy River Regional Laboratory Program (Ontario, Canada)
Kimball Medical Center (NJ)
King's Daughter Medical Center (KY)
Klinični Center (Slovenia)
LabCorp (NC)
Laboratoire de Santé Publique du Quebec (Canada)
Laboratorio Dr. Echevarne (Spain)
Laboratório Fleury S/C Ltda. (Brazil)
Laboratory Corporation of America (NJ)
LAC and USC Healthcare Network (CA)
Lakeland Regional Medical Center (FL)
Lancaster General Hospital (PA)
Langley Air Force Base (VA)
LeBonheur Children's Medical Center (TN)
L'Hotel-Dieu de Quebec (Canada)
Libero Instituto Univ. Campus BioMedico (Italy)
Lourdes Health System (NJ)
Louisiana State University Medical Center
Maccabi Medical Care and Health Fund (Israel)
Malcolm Grow USAF Medical Center (MD)
Martin Luther King/Drew Medical Center (CA)
Massachusetts General Hospital (Microbiology Laboratory)
MDS Metro Laboratory Services (Burnaby, BC, Canada)
Medical College of Virginia Hospital
Medicare/Medicaid Certification, State of North Carolina
Memorial Medical Center (IL)
Memorial Medical Center (LA) Jefferson Davis Hwy
Memorial Medical Center (LA) Napoleon Avenue
Mercy Medical Center (IA)
Methodist Hospital (TX)
MetroHealth Medical Center (OH)
Michigan Department of Community Health
Mississippi Baptist Medical Center
Monte Tabor – Centro Italo - Brazileiro de Promocao (Brazil)
Montreal Children's Hospital (Canada)
Montreal General Hospital (Canada)
National Institutes of Health (MD)
National University Hospital (Singapore)
Naval Hospital – Corpus Christi (TX)
Nebraska Health System
New Britain General Hospital (CT)
New England Fertility Institute (CT)
New England Medical Center (MA)
New Mexico VA Health Care System
New York University Medical Center
NorDx (ME)
North Carolina State Laboratory of Public Health
North Shore – Long Island Jewish Health System Laboratories (NY)
North Shore University Hospital (NY)
Northwestern Memorial Hospital (IL)
O.L. Vrouwziekenhuis (Belgium)
Ordre professionnel des technologists médicaux du Québec
Ospedali Riuniti (Italy)
The Ottawa Hospital (Ottawa, ON, Canada)
OU Medical Center (OK)
Pathology Associates Medical Laboratories (WA)
The Permanente Medical Group (CA)
Piedmont Hospital (GA)
Pocono Medical Center (PA)
Presbyterian Hospital of Dallas (TX)
Providence Health Care
Queen Elizabeth Hospital (Prince Edward Island, Canada)
Queensland Health Pathology Services (Australia)
Quest Diagnostics Incorporated (CA)
Quintiles Laboratories, Ltd. (GA)
Regions Hospital
Reid Hospital & Health Care Services (IN)
Research Medical Center (MO)
Rex Healthcare (NC)
Rhode Island Department of Health Laboratories
Riverside Medical Center (IL)
Riyadh Armed Forces Hospital (Saudi Arabia)
Robert Wood Johnson University Hospital (NJ)
Royal Columbian Hospital (New Westminster, BC, Canada)
Saad Specialist Hospital (Saudi Arabia)
Sacred Heart Hospital (MD)
Saint Mary's Regional Medical Center (NV)
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 Regional Hospital (St. John, NB, Canada)
St. John's Hospital & Health Center (CA)
St. Joseph Hospital (NE)
St. Joseph's Hospital – Marshfield Clinic (WI)
St. Joseph's Medical Center (NY)
St. Jude Children's Research Hospital (TN)
St. Luke's Regional Medical Center (IA)
St. Mary of the Plains Hospital (TX)
St. Mary's Hospital & Medical Center (CO)
St. Michael's Hospital (Toronto, ON, Canada)
St. Vincent Medical Center (CA)
Ste. Justine Hospital (Montreal, PQ, Canada)
Salina Regional Health Center (KS)
San Francisco General Hospital (CA)
Santa Clara Valley Medical Center (CA)
Sentara Williamsburg Community Hospital (VA)
Seoul Nat'l University Hospital (Korea)
South Bend Medical Foundation (IN)
So. California Permanente Medical Group
Southwest Texas Methodist Hospital (TX)
South Western Area Pathology Service (Australia)
Southern Maine Medical Center
Spartanburg Regional Medical Center (SC)
Specialty Laboratories, Inc. (CA)
Stanford Hospital and Clinics (CA)
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)
Trident Regional Medical Center (SC)
Tripler Army Medical Center (HI)
Truman Medical Center (MO)
Tuenmun Hospital (Hong Kong)
UCLA Medical Center (CA)
UCSF Medical Center (CA)
UNC Hospitals (NC)
University College Hospital (Galway, Ireland)
University Hospital (Gent) (Belgium)
University Hospitals of Cleveland (OH)
University of Alabama-Birmingham Hospital
University of Alberta Hospitals (Canada)
University of Colorado Health Science Center
University of Chicago Hospitals (IL)
University of Illinois Medical Center
University of the Ryukyus (Japan)
University of Virginia Medical Center
UZ-KUL Medical Center (Belgium)
VA (Hines) Medical Center
VA (Kansas City) Medical Center (MO)
VA (San Diego) Medical Center (CA)
VA (Tuskegee) Medical Center (AL)
Vejle Hospital (Denmark)
Virginia Department of Health
ViroMed Laboratories (MN)
Warren Hospital (NJ)
Washington Adventist Hospital (MD)
Washoe Medical Center Laboratory (NV)
Waterford Regional Hospital (Ireland)
Wellstar Health Systems (GA)
West Jefferson Medical Center (LA)
West Shore Medical Center (MI)
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 1-56238-498-8