CLINICAL AND LABORATORY STANDARDS INSTITUTE®

2nd Edition

EP06

Evaluation of Linearity of Quantitative Measurement Procedures

This guideline provides information for characterizing the linearity interval of a measurement procedure, validating a linearity interval claim (to be performed by the manufacturer), and verifying an established linearity interval claim (to be performed by the end user).

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

EP06-Ed2
November 2020
Replaces EP06-A

Evaluation of Linearity of Quantitative Measurement Procedures

Robert J. McEnroe, PhD
A. Paul Durham, MA
Marina V. Kondratovich, PhD
Jesper V. Johansen, PhD
Patrick G. Meyers, MS, CQE
Rhona J. Souers, MS
Jeffrey E. Vaks, PhD

Abstract

Clinical and Laboratory Standards Institute guideline EP06—*Evaluation of Linearity of Quantitative Measurement Procedures* is intended to provide both manufacturers and users of quantitative measurement procedures with an economical and user-friendly method of validating and verifying the linearity interval. This guideline also can be used to determine the extent to which a quantitative measurement procedure meets medical requirements or the manufacturer's linearity interval claims.

Clinical and Laboratory Standards Institute (CLSI). *Evaluation of Linearity of Quantitative Measurement Procedures.* 2nd ed. CLSI guideline EP06 (ISBN 978-1-68440-096-6 [Print]; ISBN 978-1-68440-097-3 [Electronic]). Clinical and Laboratory Standards Institute, USA, 2020.

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

Suggested Citation

CLSI. *Evaluation of Linearity of Quantitative Measurement Procedures.* 2nd ed. CLSI guideline EP06. Clinical and Laboratory Standards Institute; 2020.

Previous Editions:
October 1986, December 2001, April 2003

EP06-Ed2

ISBN 978-1-68440-096-6 (Print)

ISBN 978-1-68440-097-3 (Electronic)

ISSN 1558-6502 (Print)

ISSN 2162-2914 (Electronic)

Volume 40, Number 16

Contents

Contents (Continued)

Foreword

A measurement procedure is **linear throughout a given interval** when, in that interval, the measured results "on average" (ie, abstracting from imprecision) are **proportional** to the measurand's true quantity values, meaning that the measurand results agree with the true values up to a constant multiplicative factor:

$$\text{Measured value} = k(\text{True value}) \ (k > 0) \quad (1)$$

A measurement procedure is **linear** (without additional qualification) when the procedure is linear throughout its stated analytical measuring interval. Thus, for example, in patient monitoring, when a measurand's true value doubles or decreases by 15% from one sample to the next, results obtained using a measurement procedure demonstrated to be linear can be expected (within limits determined largely by imprecision) to respectively double or decrease by 15%, although the procedure might exhibit systematic proportional bias relative to the measurand's true quantity values.

This characterization of linearity applies not only to measurement procedures that report results in concentration units (eg, nmol/L, ng/dL, µIU/mL), but also to those reporting enzyme activity, blood cell counts, etc. (For brevity, this guideline is written as if all such assays report in concentration units.) However, some tests reporting on a continuous scale, such as tests measuring specific patient (auto)antibodies, cannot be expected to show linear behavior for all patient samples. Moreover, the characterization is consistent with the use of "linear" and cognate terms in clinical chemistry as applied to conventional linearity-under-dilution studies. These studies typically involve preparing a spectrum of mixtures by combining a high-concentration sample with a measurand-free sample (or diluent), generating and averaging replicate measurement results for each mixture, and finally regressing these results vs the values expected from the high sample proportion (ie, relative volume) represented in each mixture. Success is demonstrated when, analytically and/or graphically, the paired values (ie, observed and expected results) all closely approximate a straight-line trajectory passing through the origin (0,0), making appropriate allowance for the measurement procedure's imprecision, the experiment's size, and clinically acceptable measurand- and concentration-specific deviations from the line.

The approach advocated in this edition of EP06, as well as previous editions, can be regarded as refinements of this conventional study with respect to design, analysis, and interpretation.

Overview of Changes

This guideline replaces the previous edition of the approved guideline, EP06-A, published in 2003.

The first edition, EP06-P, published in October 1986, relied on fitting a straight line to measurements of five equally spaced samples, four replicates each, judging linearity by a goodness-of-fit test based on comparing dispersion around the regression line with the repeatability (ie, within-run imprecision) exhibited in the experiment. Unfortunately, this statistical test puts measurement procedures with excellent repeatability at risk of inappropriately failing. Conversely, it might fail to identify nonlinearity in measurement procedures with very poor repeatability.

To rectify this shortcoming, the second edition, EP06-P2, published in December 2001, and the first approved guideline, EP06-A, published in April 2003, adopted a different and computationally more complex statistical test for linearity. EP06-A called for fitting not only first-order but also second- and third-order polynomials (ie, linear, quadratic, and cubic models) to the data, judging the measurement procedure to be linear if, by internal statistical criteria, the first-order fit is best. In effect, EP06-A asked whether the trajectory of experimental results had a shape more closely resembling a straight line rather than a parabolic or sigmoidal curve. Unfortunately, this method placed no restriction on the trajectory's orientation. EP06-A, unlike major publications cited therein, was not sufficiently clear that, with suitable

allowance for random error, the trajectory should be aligned with the origin. (Intuitively, for example, a measurement procedure exhibiting little or no decrease in measured results under progressive dilutions, such as so-called "analog" procedures for free thyroxine, is not considered linear even when the trajectory of results approximates a straight-line segment.)

This edition of EP06 builds on the previous editions, introducing several important refinements, including:

- The discussion of dilution schemes, designed to minimize errors in preparing the test panels, has been extended. There is no longer any suggestion that samples need to be equally spaced. This guideline encourages judicious interpolation of additional mixtures to improve coverage of concentration gaps between calibrators, as well as concentrations important for decision-making or monitoring.
- Like EP06-A, this edition emphasizes that suitable visualizations of the study data are important, and many examples are provided.
- Consistent with other CLSI method evaluation guidelines, this guideline calls for judging results in terms of the clinical acceptability of deviations (ie, deviations from linearity at each of the sample concentrations), as opposed to a global pass-or-fail assessment based solely on internal statistical criteria. This point of view makes this guideline's approach more relevant to clinical practice and more informative as to the location, magnitude, and significance of any deviations from linearity.
- Chapter 3 is devoted to validating linearity (intended for manufacturers and developers), and Chapter 4 covers verifying (ie, spot-checking) linearity (intended for end-user laboratories).
- Two study designs are discussed: one study design includes a high sample (whose concentration is known to exceed the procedure's analytical measuring interval) and a measurand-free sample. The other study design includes high and low samples with known concentrations or a known concentration ratio. These designs serve different purposes, have different limitations, and use somewhat different data analyses.
- Computationally, this edition's approach is simpler than that of EP06-A, insofar as fitting second- and third-order polynomials is no longer included for validating or verifying linearity (although developers might find such analysis informative). Conversely, weighted first-order regression analysis is recommended under appropriate circumstances to limit the risk of failure due to chance. Advice is provided on determining adequate sample-specific weights in the absence of a precision profile.
- The importance of stating a performance claim is emphasized.

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

KEY WORDS

Linearity
Measured values
Measurement error
Proportionality
Weighted linear regression

Chapter 1

Introduction

This chapter includes:

- Guideline's scope and applicable exclusions
- Background information pertinent to the guideline's content
- Standard precautions information
- Terminology information, including:
  - Terms and definitions used in the guideline
  - Abbreviations and acronyms used in the guideline

Evaluation of Linearity of Quantitative Measurement Procedures

1 Introduction

1.1 Scope

This guideline provides recommendations for designing, analyzing, and interpreting linearity studies for quantitative measurement procedures. This guideline is intended for manufacturers and developers seeking to validate the linearity of a measurement procedure throughout a stated concentration interval, especially the interval that includes the measurement procedure's lower limit of quantitation (LLoQ) and upper limit of quantitation (ULoQ). It is also intended for laboratorians who verify the linearity of a measurement procedure and for regulatory agencies responsible for overseeing *in vitro* diagnostic (IVD) manufacturers or end-user laboratories.

This guideline does not include information on linearity issues encountered during the measurement procedure development phase, such as efficiently identifying the widest possible interval for a linearity claim or selecting calibration points, although the experimental design and data analysis principles described herein can be of value during that phase.

Before the laboratory begins formal linearity verification studies, the measurement procedure's intended analytical measuring interval claim should already have been determined based on the results of linearity, precision, and other studies that have been evaluated using a clinically informed error budget for imprecision, bias, etc.

1.2 Background

EP06 is one of the CLSI method evaluation documents, which provide guidance on experimental evaluation of quantitative measurement procedures. These documents describe studies covering, eg, precision (see CLSI documents EP05[1] and EP15[2]), measurement procedure comparison and bias (see CLSI document EP09[3]), recovery (see CLSI document EP15[2]), and limits of quantitation (see CLSI document EP17[4]). EP06 is devoted to linearity studies.

The recommendations in this guideline differ depending on whether the study is intended to validate a measurement procedure's linearity or merely to verify it and also on whether fully commutable, measurand-free material is available for use as a diluent.

For verification, practical considerations may necessitate a smaller, less rigorous study than would be required to validate performance claims for regulatory purposes. For example, compared with validation, verification may involve fewer samples (ie, mixtures, dilutions), fewer replicates, and often, for at least two reasons, sample concentrations that span only a large segment of the measurement procedure's stated analytical measuring interval. First, owing to software constraints, end users might not be able to generate explicit numerical results for samples with concentrations beyond the upper and/or lower limits of the procedure's analytical measuring interval. Moreover, owing to the procedure's inherent imprecision, laboratories might not be able to generate consistent results for samples very close to (but within) these limits. Second, to accommodate multiple commercial measurement procedures for a given measurand that differ in their stated analytical measuring intervals, third-party providers of samples for linearity (or calibration) verification studies sometimes restrict the samples' concentration span for that measurand to an interval deemed medically essential for any such procedure.

The linearity of the measurement procedure can be only demonstrated for a given interval [LLLI, ULLI] with a calculated deviation from linearity.

2.5 Linearity vs Modeling of the Calibration Relationship

A measurement procedure's linearity refers to the proportional relationship between the reported sample concentrations and the true concentrations. Sometimes the calibration relationship between the calibrator concentration (which might be not traceable to international standards) and the measurement procedure response (ie, an instrument signal) needs to be modeled. In most measurement procedures, the sample and calibrator matrixes are not identical. A calibration model is usually only an approximation of the true calibration relationship. Therefore, it can be desirable to explore calibration model options and select a calibration model that supports good measurement procedure properties. For example, in some cases it might be efficient to assume a straight-line relationship, when the true relationship has slight curvature. Because of the efficiencies provided when a calibration relationship can be represented by a straight line (ie, only two calibrators are needed), "modeling a calibration relationship" is sometimes referred to as "calibration linearity." It should not be confused with the measurement procedure's linearity.

In a measurement procedure with a linear calibration relationship $Y = f(X) = AX + B$ (where A and B are parameters of a straight line), large differences between the working and the true calibration relationships can result in unacceptably large deviations from linearity. Conversely, a measurement procedure with a nonlinear calibration relationship, when properly calibrated, can have acceptable deviation from linearity.

The terms "measurement procedure linearity" and "calibration linearity" should not be confused. "Calibration linearity" is not the subject of this guideline.

NOTE: In the following subchapters, "sample HIGH" refers to a sample with a high measurand level, "sample Blank" refers to a sample with zero measurand, and "sample LOW" refers to a sample with a low (but nonzero) measurand level. "$Value_{MP}$" denotes the average value of a large number of results produced by a measurement procedure, whereas "$Value_{TRUE}$" denotes the true value of the measurand. The letter "S" denotes a sample.

2.6 Relationship Between Linearity and Trueness

When the linearity of a measurement procedure is checked, only the **relationships** of the true measurand values between different samples need to be known. It is not necessary to know the absolute true measurand values. Practically, one of the ways to check linearity is to conduct an experiment using samples for which the relationship between true measurand values is known by formulation (eg, linearity study samples are prepared by combining sample HIGH with sample Blank in known proportions). The values of sample HIGH and sample Blank might or might not be known by measuring them with a reference measurement procedure:

- **Reference measurement procedure does not exist:** When a reference measurement procedure for evaluating a measurement procedure's bias does not exist (or is not easily available), linearity interval evaluation for the measurement procedure is especially important, because it evaluates how well the measurement procedure is calibrated with the measurement procedure calibrators. The results could identify certain types of commutability issues between patient samples and the measurement procedure calibrators. The linearity study does not provide information about trueness of the measurement procedure. However, it does

3.6.2.2 Precision Profile

For preliminary evaluation of whether the repeatability in the linearity study is acceptable and for determination of the weights to be used in the linear regression, the precision profile from the linearity study is determined and compared for consistency with the precision profile from the precision studies specified in CLSI document EP05.[1] See Subchapter 3.6.3 for an example of preliminary evaluation of repeatability and Subchapter 3.6.4.2 for information on WLS linear regression.

The linearity data might need to be divided into subintervals for analysis, depending on the number of replicates:

- **Four or more replicates:** In general, each level S_i should be measured with at least four replicates for each sample level, when possible. Four or more replicates are sufficient to determine the precision profile as % CV (and SD) vs measured concentration (ie, mean values of R replicates).
- **Three replicates:** When three replicates are tested for each level in the linearity study, the linearity interval is divided into four subintervals such that each subinterval has two to three levels. For each of these subintervals' levels, the pooled variances of the replicates and the mean value of all replicates are estimated from the samples that belong to this subinterval (see example 1 in Subchapter 3.6.3).
- **Two replicates:** When two replicates are tested for each level in the linearity study, the linearity interval is divided into three subintervals such that each subinterval has at least three levels. For each of these subintervals' levels, the pooled variances of the replicates and the mean value of all replicates are estimated from the samples that belong to this subinterval (see example 1 in Subchapter 3.6.3).

3.6.3 Example 1

For example 1, when a subinterval has three levels with SD_1, SD_2, and SD_3 of repeatability corresponding to R_1, R_2, and R_3 replicates, the pooled variance is:

$$Var = \frac{(R_1 - 1)SD_1^2 + (R_2 - 1)SD_2^2 + (R_3 - 1)SD_3^2}{(R_1 - 1) + (R_2 - 1) + (R_3 - 1)} \quad (30)$$

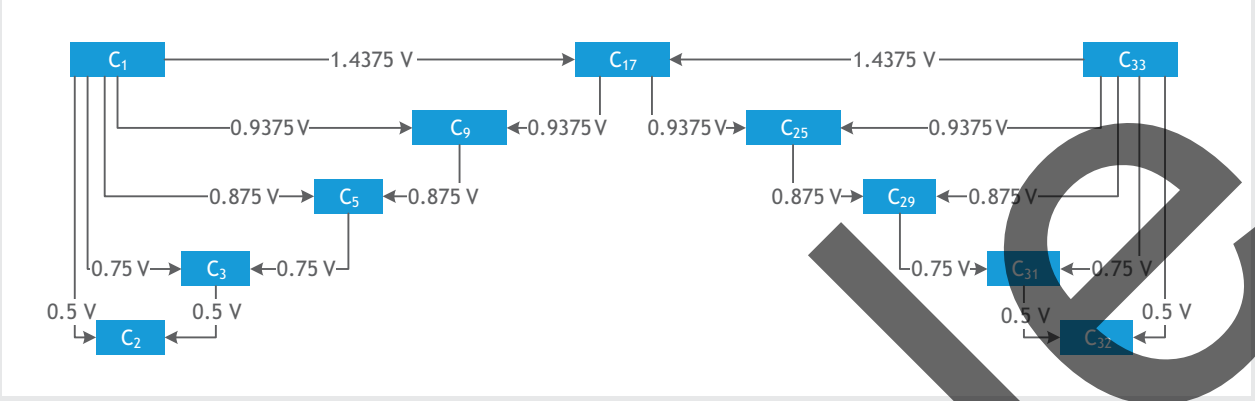
When all levels are measured with the same number of replicates, R:

$$Var = \frac{SD_1^2 + SD_2^2 + SD_3^2}{3} \quad (31)$$

Components of variance relevant to the linearity study are obtained from the precision study to form a total variance, which is used to determine limits for imprecision in the linearity study. According to CLSI document EP05,[1] components of variance such as within-run, between-run, and between-day can be extracted from the precision study. These variance components might not all be present in the linearity study. For example, when only one run is used to obtain all M • R measurements, the limits of imprecision are based on the repeatability (ie, within-run) component. When the M • R measurements are obtained from different runs, the limits of imprecision are based on within-run and between-run components.

The precision profile of the pooled SD in the linearity study is visually compared for consistency with the precision profile from the precision studies specified in CLSI documents EP05[1] and EP17.[4] When the precision profile in the linearity study is not consistent with the precision profile from the precision studies, the inconsistency should be investigated (see CLSI document EP05[1]).

Appendix C. (Continued)



Symbols: *C*, concentration; *V*, volume.

Figure C5. Flow Chart for Preparing 11 Concentration Levels With Relative Spacing Symmetrically Decreasing From the Middle as Series 8, 4, 2, 1

NOTE: Each of the 11 final volumes, at each concentration, is V. The initial volumes of concentrations C_1 and C_{33} are 5.5 V each.

The mixing schemes in Figures C2 to C5 enable smaller concentration spacings near the expected limits of the linearity interval (ie, 6.25% of the entire interval for the 17-level scheme and 3.125% for the 33-level scheme) rather than around the middle of the interval. In some cases, tightening the claimed linearity interval by removing a level at the extreme helps meet the acceptance criteria without collecting new data, while sacrificing little in terms of the width of the interval. Other numbers of levels and concentration spacings can be achieved using the equal-volume mixing scheme.

Volumetric errors accumulate with each step of an equal-volume mixing scheme. When there are more than five mixing steps, equal-volume mixing schemes lose the advantage of smaller volumetric errors vs the proportional mixing scheme described in section C2.[2]

C2 Proportional Mixing of Low and High Concentrations

As shown in Table C1, with the proportional mixing scheme, linearity panel members can have any arbitrary concentrations between the high and low concentrations used for mixing. A panel member with concentration C_i can be obtained by mixing f_i proportion of concentration C_1 (low) with $(1 - f_i)$ proportion of concentration C_N (high), the total number of levels being N:

$$C_i = f_i C_1 + (1 - f_i) C_N \quad \text{(C11)}$$

Solving equation (C11) for f_i yields:

$$f_i = \frac{C_N - C_i}{C_N - C_1} \quad \text{(C12)}$$

Appendix G. An Example of a Linearity Study With Two Replicates

Abbreviations for Appendix G

% deviation	deviation from linearity expressed as a percentage
ADL	allowable deviation from linearity
LLoD	lower limit of detection
SD	standard deviation
WLS	weighted least squares

Symbols for Appendix G

A	slope
E	expected value
W	weight
Y	measured value for panel member

NOTE: For calculations, rounding should only be performed on the final results. When performing the calculations included in this guideline, users might obtain values that differ slightly from those shown, depending on the statistical application used.

The proposed linearity interval is 55 to 500 U/L and the allowable deviation from linearity (ADL) is ± 10%. Ten samples are prepared by diluting sample HIGH, which is above 500 U/L, with sample Blank, which has a value of zero. Each sample in the linearity study is tested twice, and all 20 replicates are obtained in a single run. Results are shown in Table G1.

Table G1. Linearity Study Results

Dilution	RC	Replicate 1	Replicate 2	Measured Value (mean)	SD	% CV
1	1	525.0	533.0	529.00	5.657	1.07
2	0.9	483.3	510.0	496.62	18.880	3.80
3	0.8	453.8	482.6	468.20	20.365	4.36
4	0.7	383.5	399.4	391.46	11.243	2.87
5	0.6	346.7	353.6	350.20	4.879	1.40
6	0.5	287.6	296.5	292.04	6.293	2.16
7	0.4	229.2	237.1	233.15	5.586	2.41
8	0.3	171.7	172.2	171.93	0.354	0.18
9	0.2	110.0	112.6	111.32	1.838	1.61
10	0.1	51.0	51.5	51.25	0.354	0.73

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

Figure G1 presents individual replicates vs mean measured value. A visual assessment of the data shows no apparent outliers.

Related CLSI Reference Materials[a]

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

EP07 **Interference Testing in Clinical Chemistry. 3rd ed., 2018.** This guideline provides background information, guidance, and experimental procedures for investigating, identifying, and characterizing the effects of interferents on clinical chemistry test results.

EP09 **Measurement Procedure Comparison and Bias Estimation Using Patient Samples. 3rd ed., 2018.** This guideline covers the design of measurement procedure comparison experiments using patient samples and subsequent data analysis techniques used to determine the bias between two *in vitro* diagnostic measurement procedures.

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

EP17 **Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures. 2nd ed., 2012.** This document provides guidance for evaluation and documentation of the detection capability of clinical laboratory measurement procedures (ie, limits of blank, detection, and quantitation), for verification of manufacturers' detection capability claims, and for the proper use and interpretation of different detection capability estimates.

EP34 **Establishing and Verifying an Extended Measuring Interval Through Specimen Dilution and Spiking. 1st ed., 2018.** It is often medically necessary to provide results for specimens with concentrations above the analytical measuring interval of an *in vitro* diagnostic measurement procedure. This guideline helps manufacturers and laboratory scientists with establishing, validating, or verifying a dilution scheme that will provide an extended measuring interval for such specimens.

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

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

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

PRINT ISBN 978-1-68440-096-6

ELECTRONIC ISBN 978-1-68440-097-3

EP06-Ed2