CLINICAL AND LABORATORY STANDARDS INSTITUTE®

3rd Edition

EP09c

Measurement Procedure Comparison and Bias Estimation Using Patient Samples

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.

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

EP09c, 3rd ed.
June 2018
Replaces EP09-A2-IR and EP09-A3

Measurement Procedure Comparison and Bias Estimation Using Patient Samples

Jeffrey R. Budd, PhD
A. Paul Durham, MA
Thomas E. Gwise, PhD
Douglas M. Hawkins, PhD
Mark Holland, PhD
Beimar Iriarte, MS
Anders Kallner, MD, PhD
Kristian Linnet, MD, PhD
Robert Magari, PhD
Jeffrey E. Vaks, PhD

Abstract

Clinical and Laboratory Standards Institute guideline EP09—*Measurement Procedure Comparison and Bias Estimation Using Patient Samples* is written for laboratorians and manufacturers. It describes procedures for determining the bias between two measurement procedures, and it identifies factors for consideration when designing and analyzing a measurement procedure comparison experiment using patient samples. An overview of the measurement procedure comparison experiment includes considerations for both manufacturers and laboratorians. Details on how to create difference and scatter plots for visual inspection of the data are provided. Once the data are characterized, various methods are introduced for quantifying the relationship between two measurement procedures, including bias estimates and regression techniques. The final chapter contains recommendations for manufacturers' evaluation of bias and statement format for bias claims.

Clinical and Laboratory Standards Institute (CLSI). *Measurement Procedure Comparison and Bias Estimation Using Patient Samples*. 3rd ed. CLSI guideline EP09c (ISBN 978-1-68440-006-5 [Print]; ISBN 978-1-68440-007-2 [Electronic]). Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087 USA, 2018.

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

Suggested Citation*

CLSI. *Measurement Procedure Comparison and Bias Estimation Using Patient Samples*. 3rd ed. CLSI guideline EP09c. Wayne, PA: Clinical and Laboratory Standards Institute; 2018.

Previous Editions:
January 1986, April 1993, December 1995, September 2002, July 2010, August 2013

ISBN 978-1-68440-006-5 (Print)
ISBN 978-1-68440-007-2 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic) Volume 38, Number 12

* EP09 was corrected in June 2018, and the code was revised to EP09c.

Contents

Contents (Continued)

Foreword

Measurement procedure comparison is one of the most common techniques used by both manufacturers and medical laboratorians to estimate the bias of an *in vitro* diagnostic (IVD) measurement procedure relative to a comparator. It involves the comparison of results from patient samples from two measurement procedures intended to measure the same component (eg, measurand concentration) with the key determination being the estimate of bias between them.

A number of different scenarios exist in which measurement procedure comparison studies are indicated. For both the manufacturer and the medical laboratorian, the ideal scenario is the comparison of a candidate measurement procedure to a generally accepted standard or reference measurement procedure. In the case of a manufacturer, this involves the establishment and perhaps validation of performance claims for bias. In the case of a laboratorian, it involves introducing a measurement procedure into the laboratory, including verification of its manufacturer's claims (specifications). The scope of the experimental and data-handling procedures for these two purposes differs. In either case the assumption that the reference measurement procedure provides "true" values means that bias (systematic measurement error) is estimated.

Quite commonly, however, there is no standard or reference measurement procedure. The manufacturer instead compares a candidate measurement procedure to the most appropriate measurement procedure currently available. The laboratorian usually compares the candidate and an available procedure. Then, there might not be a "true" value and the "difference," rather than the "bias," is estimated.

Given the variety of performance characteristics of IVD measurement procedures, a single experimental design is not appropriate for all types of laboratory and manufacturer measurement procedure comparisons. Therefore, performance characteristics such as measuring interval and precision profile are taken into account in structuring an experiment for comparing two measurement procedures. Multiple worked examples are presented.

This guideline is intended to promote effective and correct data analysis and reporting using standard experimental and statistical methods.

Manufacturers of medical laboratory measurement procedures or devices should use this guideline to establish and standardize their bias performance claims. Many different forms have been used for such claims, and they have not always been sufficiently specific to allow user verification.

Overview of Changes

This guideline replaces the previous edition of the approved guideline, EP09-A2-IR, published in 2010. Several changes were made in this edition, including:

- Broader coverage of measurement procedure comparison applications
- More reasons for comparisons based on patient samples (factor comparisons [eg, sample tube types])
- Visualization/exploration of data using difference plots
- Regression descriptions including weighted options, Deming, and Passing-Bablok techniques
- Measurement of bias using difference plots
- Measurement of bias at clinical decision points
- Computation of confidence intervals for all parameters
- Outlier detection using extreme studentized deviate
- Relocation of most of the detailed mathematical descriptions to the appendixes

This guideline was corrected in 2018 and replaces the original third edition of the approved guideline, EP09-A3, published in 2013. Corrections were made as follows:

- Reorganizing the content to emphasize the process of performing a measurement procedure comparison
- Clearly specifying that manufacturers should use regression analysis to characterize bias
- Adding information on using precision profile information in performing Deming regressions
- Adding more information on determining confidence intervals for bias estimates at specified concentrations using regression fits
- Making corrections to the description of the Passing-Bablok regression technique
- Adding a detailed description of the bootstrap iterative technique for bias estimation
- Correcting minor miscellaneous errors in equations

> **NOTE:** Due to the complex nature of the calculations in this guideline, it is recommended that the user have access to a computer and statistical software.

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

Alternative regression methods, bias, evaluation protocol, experimental design, linear regression, measurement procedure comparison, outliers, quality control, residuals

Measurement Procedure Comparison and Bias Estimation Using Patient Samples

Chapter 1: Introduction

This chapter includes:

- Guideline's scope and applicable exclusions
- Standard precautions information
- "Note on Terminology" that highlights particular use and/or variation in use of terms and/or definitions
- Terms and definitions used in the guideline
- Abbreviations and acronyms used in the guideline
- Symbols used in the guideline

1.1 Scope

This guideline provides recommendations for designing an experiment and selecting methods to quantify systematic measurement error (bias or difference) between measurement procedures based on comparing patient samples. It provides both difference plot and regression procedures to determine the relationship between two measurement procedures either across their measuring intervals or at selected concentrations. Intended users of this guideline are manufacturers of *in vitro* diagnostic (IVD) reagents—which includes those who create laboratory-developed tests—as well as regulatory authorities and medical laboratory personnel.

This guideline is for use with measurement procedures that provide quantitative numerical results. This guideline is not intended for use with ordinal IVD examinations, commonly referred to as qualitative procedures (see CLSI document EP12[1]). This guideline is not intended to provide information on evaluation of random error (see CLSI documents EP05[2] and EP15[3]) or to determine the total error inherent in a comparison of measurement procedures (see CLSI document EP21[4]). It is not intended to measure the variability of multiple replicates collected during the measurement of a sample, nor is it intended to measure the bias of individual measurements such as those resulting from sample interference (as covered in CLSI document EP07[5]).

1.2 Standard Precautions

Because it is often impossible to know what isolates or specimens might be infectious, all patient and laboratory specimens are treated as infectious and handled according to "standard precautions." Standard precautions are guidelines that combine the major features of "universal precautions and body substance isolation" practices. Standard precautions cover the transmission of all known infectious agents and thus are more comprehensive than universal precautions, which are intended to apply only to transmission of bloodborne pathogens. Published guidelines are available that discuss the daily operations of diagnostic medicine in humans and animals while encouraging a culture of safety in the laboratory.[6] For specific precautions for preventing the laboratory transmission of all known infectious agents from laboratory instruments and materials and for recommendations for the management of exposure to all known infectious diseases, refer to CLSI document M29.[7]

1.3 Terminology

1.3.1 A Note on Terminology

CLSI, as a global leader in standardization, is firmly committed to achieving global harmonization whenever possible. Harmonization is a process of recognizing, understanding, and explaining differences while taking steps to achieve worldwide uniformity. CLSI recognizes that medical conventions in the global metrological community have evolved differently in different countries and regions and that legally required use of terms, regional usage, and different consensus timelines are all important considerations in the harmonization process. CLSI recognizes its important role in these efforts, and its consensus process focuses on harmonization of terms to facilitate the global application of standards and guidelines.

Essentially, new documents must adhere to the latest international vocabulary of metrology[8] whenever an ambiguity in the interpretation or understanding of terms occurs. Recently, many definitions have become more explicit and understandable, but the international language is difficult and compact. The international vocabulary of metrology deals with general metrology and terminology that should be useful for most disciplines that measure quantities.

The understanding of a few terms has changed during the last decade as the concepts have developed. Particularly, *trueness* (measurement trueness) is defined as expressing the closeness of agreement between the average of an infinite number of replicate measurements and a reference value; and *precision* (measurement precision) is defined as closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions. Consequently, *accuracy* (measurement accuracy) is the closeness of agreement between a measured value and a true quantity value of a measurand. Thus, this concept comprises both trueness and precision, and applies to a single result. *Measuring interval* has replaced *reportable range* when referring to "a set of values of a measurand for which the error of a measuring instrument (test) is intended to lie within specified limits." An *interval* [a;b] is delineated by two limits a and b ($b > a$), whereas a *range* (r[a;b]) is expressed as the difference between b and a ($b - a$). Thus, the range of the interval [a;b] is the difference ($b - a$) that is denoted by r[a;b].

The term *measurand* is used when referring to the quantity intended to be measured instead of *analyte* (component represented in the name of a measurable quantity). The term *measurement procedure* replaces *analytical method* and *assay* for a set of operations, used in the performance of particular measurements according to a given method.

Verification focuses on whether specifications of a measurement procedure can be achieved, whereas *validation* verifies that the procedure is fit for purpose.

NOTE: Mandates are generally reserved for CLSI standards but are occasionally allowed in CLSI guidelines. In CLSI guidelines, use of the term "must" is either 1) based on a requirement or 2) indicative of a necessary step to ensure patient safety or proper fulfillment of a procedure. The working group evaluated use of the term "must" and deemed it appropriate.

1.3.2 Definitions

accuracy (of measurement) – closeness of agreement between a measured quantity value and a true quantity value of a measurand[8]; **NOTE 1:** The concept "measurement accuracy" is not a quantity and is not given a numerical quantity value. A measurement is said to be more accurate when it offers a smaller measurement error[8]; **NOTE 2:** The term "measurement accuracy" should not be used for "measurement trueness," and the term "measurement precision" should not be used for "measurement accuracy," which, however, is related to both these concepts[8]; **NOTE 3:** "Measurement accuracy" is sometimes understood as closeness of agreement between measured quantity values that are being attributed to the measurand.[8]

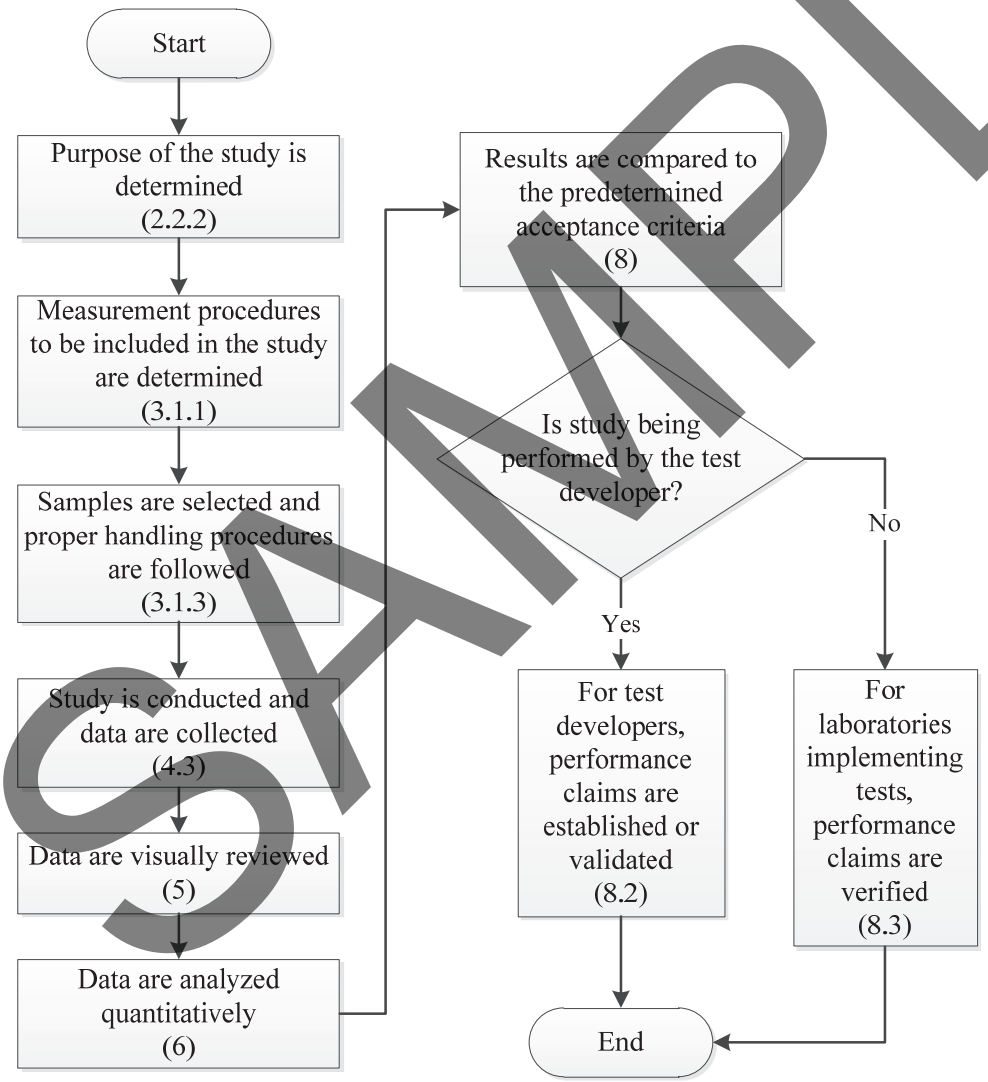
Chapter 2: Introduction to Measurement Procedure Comparison and Bias Estimation

This chapter includes:

- Process flow chart
- Introduction to measurement procedure comparison
- Overview of the study
- Purposes for performing measurement procedure comparisons

2.1 Process Flow Chart

Figure 1 shows the measurement procedure comparison process for test developers and laboratories. Applicable chapters and subchapters for each topic are included parenthetically.



* Five basic symbols are used in process flow charts: oval (signifies the beginning or end of a process), arrow (connects process activities), box (designates process activities), diamond (includes a question with alternative "Yes" and "No" responses), pentagon (signifies another process).

Figure 1. Process for Measurement Procedure Comparison*

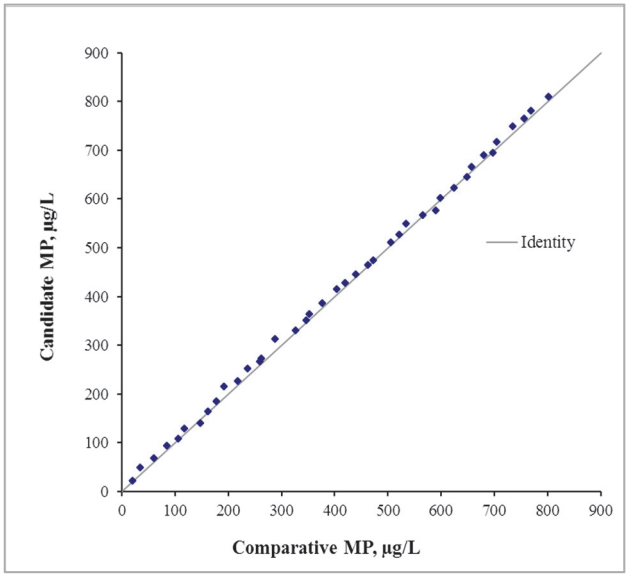


Figure 2. Scatter Plot With Constant SD (From Table D1)
Abbreviation: MP, measurement procedure.

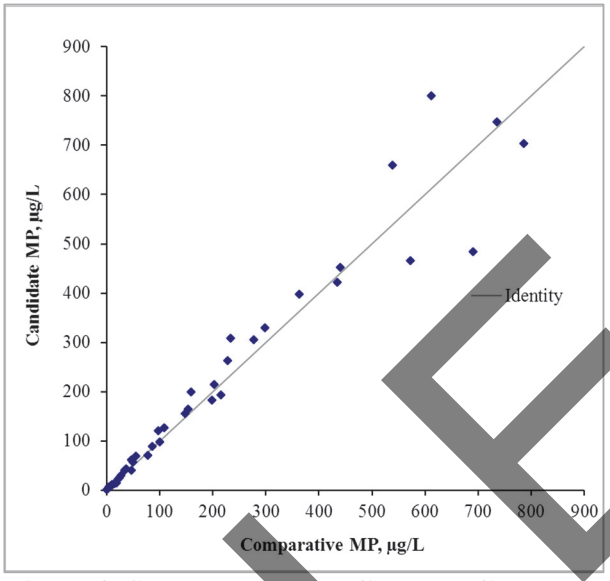


Figure 3. Scatter Plot With Constant CV (From Table D2)
Abbreviation: MP, measurement procedure.

5.3 Difference Plots

A difference plot[23] presents the results of a measurement procedure comparison study, with the measurand concentration on the horizontal axis and the difference between the candidate and the comparative measurement procedures on the vertical axis (see Figure 4). Bland-Altman[22] is an example of a difference plot. Such plots can be visually inspected to determine the underlying variability characteristics of this relationship.

The user must select from four types of difference plots based on two factors. The first factor is determined by whether the user wishes to see the comparative method as the truth against which the candidate method is compared or to see the average of the two methods as the best estimate of the true value for a sample. In the first case, the horizontal axis of the plot is the result from the comparative measurement procedure.[24] This option should be used for presenting validation data. In the second case, advocated by Bland and Altman,[25] the horizontal axis is the average of the two measurement procedures' results.

When a reference measurement procedure is the comparative measurement procedure, its results should be used on the horizontal axis. A manufacturer may wish to use the most common measurement procedure as the comparative measurement procedure. In this case, when the comparative measurement procedure is not considered a reference, the average result of the two measurement procedures (candidate and comparative) may be used on the horizontal axis for data visualization during the establishment phase of test development.

A medical laboratory may use its current measurement procedure as the comparative measurement procedure and may consider it to be a reference because the goal is to compare the known behavior of its current procedure against the unknown candidate measurement procedure. In this case, the results for the comparative measurement procedure should be used on the horizontal axis.

The second factor is whether the variability of the differences between the two measurement procedures is constant or proportional to the concentration on the horizontal axis. In the first instance, the magnitude of the difference is assumed to be essentially the same across the entire interval of concentrations (see Figure 2). In the second instance, the magnitude of the difference is assumed to be proportional to concentration (see Figure 3). Because this characteristic of the relationship might not be known beforehand, it is suggested