



January 2008

# EP12-A2

## User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline—Second Edition

This document provides a consistent approach for protocol design and data analysis when evaluating qualitative diagnostic tests. Guidance is provided for both precision and method-comparison studies.

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

# Clinical and Laboratory Standards Institute

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

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

## Consensus Process

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

## Commenting on Documents

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

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

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

## Appeals Process

If it is believed that an objection has not been adequately addressed, the process for appeals is documented in the CLSI Administrative Procedures.

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

## Get Involved—Volunteer!

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

For further information on committee participation or to submit comments, contact CLSI.

Clinical and Laboratory Standards Institute  
950 West Valley Road, Suite 2500  
Wayne, PA 19087 USA  
P: 610.688.0100  
F: 610.688.0700  
[www.clsi.org](http://www.clsi.org)  
[standard@clsi.org](mailto:standard@clsi.org)

ISBN 1-56238-654-9  
ISSN 0273-3099

EP12-A2  
Vol. 28 No. 3  
Replaces EP12-A  
Vol. 22 No. 14

---

## User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline—Second Edition

Volume 28 Number 3

Patricia E. Garrett, PhD  
Fred D. Lasky, PhD  
Kristen L. Meier, PhD

### Abstract

Clinical and Laboratory Standards Institute document EP12-A2—*User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline—Second Edition* provides the user with a consistent approach for protocol design and data analysis when evaluating qualitative diagnostic tests. Guidance is provided for both precision and method-comparison studies.

Clinical and Laboratory Standards Institute (CLSI). *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline—Second Edition*. CLSI document EP12-A2 (ISBN 1-56238-654-9). Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087 USA, 2008.

The Clinical and Laboratory Standards Institute consensus process, which is the mechanism for moving a document through two or more levels of review by the health care community, is an ongoing process. Users should expect revised editions of any given document. Because rapid changes in technology may affect the procedures, methods, and protocols in a standard or guideline, users should replace outdated editions with the current editions of CLSI documents. Current editions are listed in the CLSI catalog and posted on our website at [www.clsi.org](http://www.clsi.org). If your organization is not a member and would like to become one, and to request a copy of the catalog, contact us at: Telephone: 610.688.0100; Fax: 610.688.0700; E-Mail: [customerservice@clsi.org](mailto:customerservice@clsi.org); Website: [www.clsi.org](http://www.clsi.org).



Copyright ©2008 Clinical and Laboratory Standards Institute. Except as stated below, any reproduction of content from a CLSI copyrighted standard, guideline, companion product, or other material requires express written consent from CLSI. All rights reserved. Interested parties may send permission requests to [permissions@clsi.org](mailto:permissions@clsi.org).

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

### **Suggested Citation**

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

### **Proposed Guideline**

July 2000

### **Approved Guideline**

August 2002

### **Approved Guideline—Second Edition**

January 2008

ISBN 1-56238-654-9  
ISSN 0273-3099

## Committee Membership

### Area Committee on Evaluation Protocols

**Luann Ochs, MS**  
**Chairholder**  
**BD Diagnostics – TriPath**  
**Durham, North Carolina**

**Greg Cooper, CLS, MHA**  
**Vice-Chairholder**  
**Bio-Rad Laboratories, Inc.,**  
**QSD Division**  
**Plano, Texas**

George S. Cembrowski, MD, PhD  
University of Alberta Hospital  
Edmonton, Alberta, Canada

David L. Duewer, PhD  
National Institute of Standards and  
Technology  
Gaithersburg, Maryland

Anders Kallner, MD, PhD  
Karolinska Hospital  
Stockholm, Sweden

Kristian Linnet, MD, PhD  
University of Copenhagen  
Copenhagen, Denmark

Donald R. Parker, PhD  
Bayer HealthCare, LLC  
Mishawaka, Indiana

Lakshmi Vishnuvajjala, PhD  
FDA Ctr. for Devices/Rad. Health  
Rockville, Maryland

#### Advisors

David A. Armbruster, PhD,  
DABCC, FACB  
Abbott  
Abbott Park, Illinois

R. Neill Carey, PhD  
Peninsula Regional Medical Center  
Salisbury, Maryland

Carl C. Garber, PhD, FACB  
Quest Diagnostics, Incorporated  
Lyndhurst, New Jersey

Patricia E. Garrett, PhD  
SeraCare Life Sciences, Inc.  
Portland, Maine

Martin H. Kroll, MD  
Boston Medical Center  
Boston, Massachusetts

Jan S. Krouwer, PhD  
Krouwer Consulting  
Sherborn, Massachusetts

Jacob (Jack) B. Levine, MBA  
Siemens Medical Solutions  
Diagnostics  
Tarrytown, New York

Donald M. Powers, PhD  
Powers Consulting Services  
Pittsford, New York

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

Gian Alfredo Scassellati, PhD  
Ente Nazionale Italiano  
Di Unificazione  
Turin, Italy

Michele M. Schoonmaker, PhD  
Cepheid  
Sunnyvale, California

Daniel W. Tholen, MS  
American Association for  
Laboratory Accreditation  
Traverse City, Michigan

Jack Zakowski, PhD, FACB  
Beckman Coulter, Inc.  
Brea, California

### Working Group on Qualitative Test Performance

**Patricia E. Garrett, PhD**  
**Chairholder**  
**SeraCare Life Sciences, Inc.**  
**Portland, Maine**

Fred D. Lasky, PhD  
Genzyme Diagnostics  
Cambridge, Massachusetts

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

#### Staff

Clinical and Laboratory Standards  
Institute  
Wayne, Pennsylvania

Lois M. Schmidt, DA  
*Vice President, Standards  
Development and Marketing*

Jane M. Oates, MT(ASCP)  
*Staff Liaison*

Patrice E. Polgar  
*Project Manager*

Melissa A. Lewis  
*Editor*

### Acknowledgment

CLSI, the Area Committee on Evaluation Protocols, and the Working Group on Qualitative Test Performance gratefully acknowledge Marina Kondratovich, PhD, FDA Center for Devices/Radiological Health, who provided extensive statistical comments, insight, and expertise, and contributed to numerous discussions in revising this document; and Robert Sarber, PhD, Genzyme Diagnostics, who provided insightful statistical comments and the graphics for this consensus guideline.



**Contents**

Abstract.....i

Committee Membership..... iii

Foreword..... vii

Laboratory Error Sources and CLSI Evaluation Protocols Documents.....ix

1 Scope.....1

2 Introduction.....1

3 Standard Precautions.....2

4 Clinical Utility .....2

    4.1 Screening Tests .....3

    4.2 Diagnostic Tests.....3

    4.3 Confirmatory Tests .....3

5 Terminology.....3

    5.1 A Note on Terminology .....3

    5.2 Terms for Comparative Benchmark.....4

    5.3 Definitions .....5

6 Device Familiarization and Training .....8

    6.1 Purpose .....8

    6.2 Duration .....8

7 Evaluation Materials .....8

    7.1 Controls.....8

    7.2 Specimen Collection and Handling .....8

8 Bias and Imprecision Studies.....8

    8.1 Analyte Concentrations Near the Cutoff.....9

    8.2 Negative and Positive Controls.....11

    8.3 A Qualitative Method-Precision Experiment for Analyte Concentrations Near  
    the C<sub>50</sub> .....12

9 Comparison of Methods.....14

    9.1 Test Specimens .....15

    9.2 Number of Specimens.....16

    9.3 Duration .....16

    9.4 Inspection of Data During Collection .....16

    9.5 Discrepant Results .....16

    9.6 Reference Specimen Panels .....17

10 Data Analysis.....17

    10.1 Comparator Is Diagnostic Accuracy Criteria.....17

    10.2 Comparator Other Than Diagnostic Accuracy Criteria .....23

    10.3 Examples.....25

References.....32

Appendix. Statistical Reasoning for Precision Experiment Conclusions .....34

Summary of Consensus Comments and Working Group Responses .....40

**Contents (Continued)**

Summary of Delegate Comments and Working Group Responses .....41  
The Quality Management System Approach .....44  
Evaluation Protocols Documents, Descriptions, and Key Words.....45



## Foreword

Qualitative diagnostic tests have been used since the early days of laboratory medicine for the screening, diagnosis, and management of a variety of diseases. These tests are found in many specialties of the clinical laboratory. Method evaluation procedures for such tests are diverse, with each laboratory specialty often emphasizing different issues in both the experimental design and in the data analysis and interpretation of such studies.

There have been four key published efforts to standardize both the experimental details as well as the data analysis of qualitative information. The International Federation of Clinical Chemistry published a guideline in 1989 on protocol design and data analysis, featuring examples for urinary glucose and albumin by visually read reagent strips.<sup>1</sup> The European Committee for Clinical Laboratory Standards published a guideline in 1990 that focused on the evaluation of qualitative tests.<sup>2</sup> A very prominent work on the assessment of both quantitative and qualitative laboratory tests, *Beyond Normality: The Predictive Value and Efficiency of Medical Diagnosis*, was written by Gambino and Galen in 1975.<sup>3</sup> In addition, CLSI/NCCLS document GP10<sup>4</sup> describes the assessment of the diagnostic accuracy of a test compared to the clinical status of the patient.

The latter two publications both focus on the relation of the test result to the clinical status of the patient, for either qualitative or quantitative tests. In many laboratories, the clinical information is not readily available, so it is important that protocols for evaluations are established that enable comparison of a new test (candidate method) to other laboratory procedures, in much the same way that most method evaluation studies are performed for quantitative tests. Ideally, the comparison should be made to “diagnostic accuracy criteria” or to a so-called “gold standard.” However, comparison with a method in current use is also of interest. This guideline describes two different situations for these studies: the first is when the laboratory knows the clinical diagnosis of each specimen in the study, and the second is when the laboratory does not know the clinical diagnosis of each patient specimen. These are treated separately, to enable appropriate data analysis. Parameters such as specificity and sensitivity, and predictive value for the candidate method are estimated in the former situation, and agreement measures are estimated in the latter situation.

This guideline is intended to promote uniformity in performance assessment of qualitative tests among:

- laboratories of all types that perform qualitative tests;
- manufacturers of qualitative diagnostic tests, for describing test performance and helping customers design test performance verification studies; and
- regulatory agencies and laboratory surveyors.

During the review and revision of EP12-A, the working group added to the evaluation protocol information on the confidence limits achieved by different numbers of observations made and different results observed, allowing the user to appreciate directly the difference in confidence limits when fewer or more observations are made. In addition, some new terms were included and defined. The term “C<sub>5</sub>-C<sub>95</sub> interval” is used to describe the range of analyte concentrations around the cutoff such that observed results at concentrations *outside* this interval are consistently negative (concentrations <C<sub>5</sub>) or consistently positive (concentrations >C<sub>95</sub>). Observed results at concentrations inside this interval are not consistent due to imprecision. These concepts are explained in greater detail in Section 8.

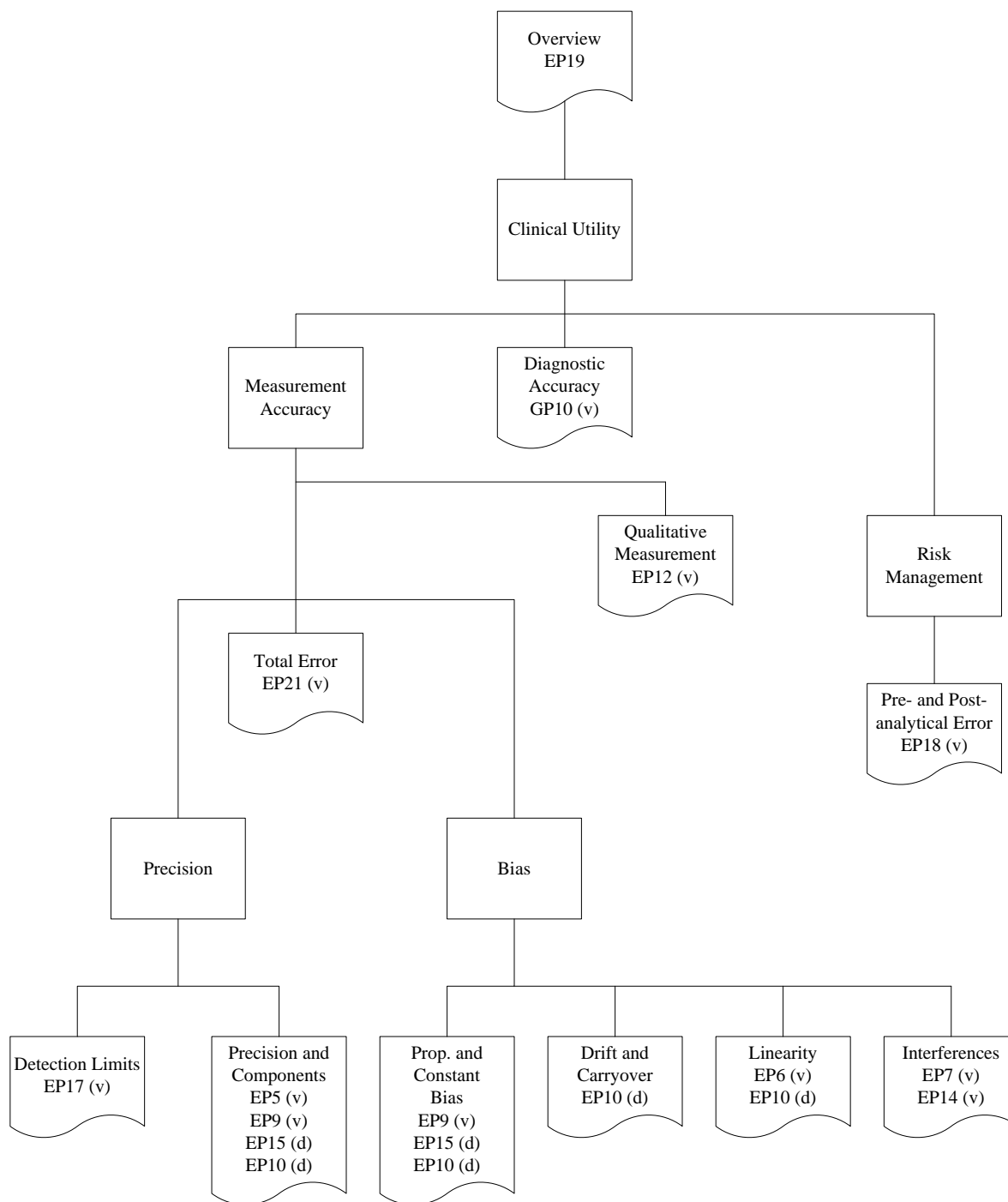
A figure was also added to illustrate which error sources the EP12 protocol can detect with respect to all error sources and other Evaluation Protocols (EP) documents (see page ix). In addition, the working group added an appendix that describes and discusses the confidence intervals used in the protocol. The

concepts are complex, and the working group encourages comments, questions, and suggestions for improvements in future revisions of the document.

**Key Words**

Analytical goals, evaluation protocol, qualitative test

## Laboratory Error Sources and CLSI Evaluation Protocols Documents



**Laboratory Error Sources and CLSI Evaluation Protocols Documents.**<sup>a</sup> This figure illustrates the relationship among parameters estimated by EP documents. Items higher up in the figure are more comprehensive whereas lower level items are more specific. Overall, the figure is much like a cause-and-effect diagram. Documents marked (d) provide guidance for demonstrating that a source of measurement inaccuracy is within acceptable limits. Documents marked (v) provide guidance for more rigorous evaluation of inaccuracy components.

<sup>a</sup> For a description of each of the documents listed, please see page 45.



## **User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline—Second Edition**

### **1 Scope**

This guideline provides protocols for the evaluation of qualitative test performance characteristics. In this document, a qualitative test is restricted to those tests that have only two possible outcomes (eg, positive/negative, present/absent, reactive/nonreactive). EP12 is written primarily for individuals and laboratories that use and evaluate such tests. These protocols are intended to help users determine test performance in their own testing environment. This guideline for qualitative test performance evaluation should help the device developer and the user to meet documentation and regulatory goals. While this document is not intended for manufacturers to establish test performance characteristics, the data analysis principles described here can be used by manufacturers.

Test methods with values that are reported as, for instance, negative, +1, +2, or +3, or as endpoint dilutions (commonly in multiples of 8, reflecting the microtiter plate format, or in multiples of 10) are often called semiquantitative. These methods are not further discussed in EP12, although if one of the values results is considered the cutoff for a positive test, the evaluation protocol recommended here could be applied to that cutoff. For instance, if a test for antibodies to the Lyme disease pathogen was reported as positive if the endpoint titer was  $\geq 1:160$ , the precision and method comparison experiments discussed below could be applied to that cutoff.

### **2 Introduction**

Qualitative tests return one of two possible responses. Method evaluation procedures for such tests are diverse, with each laboratory specialty often emphasizing different issues in the experimental design, data analysis, and interpretation of such studies. EP12 offers a defined approach to method evaluation for many qualitative tests.

Clinical laboratories develop and implement qualitative tests for a number of reasons. Laboratories should document that the test performs as intended in their facilities, by operators who are expected to use the device. Often, such demonstration is required by laboratory regulatory or accreditation bodies.

Qualitative candidate methods are diverse, employing technologies from lateral flow with visual reading, to automated nucleic acid sequencing and base calling, to microarrays. While universal evaluation guidelines may not be feasible, common features exist. For example, precision studies and comparison of methods studies with patient specimens can be used to demonstrate each type of test's performance capabilities.

In addition to technical diversity, qualitative tests may differ just by having a different cutoff or medical decision point. For example, a qualitative candidate method for screening blood donations for an infectious disease, such as human immunodeficiency virus (HIV) or hepatitis B virus (HBV), would have a cutoff chosen to ensure a high sensitivity and high negative predictive value (NPV), so there is a correspondingly high probability that an infected unit would be excluded from the blood supply. The same method could be used to make a qualitative test for diagnosis, with a cutoff that is chosen to minimize the number of false-positive and false-negative results.

When a qualitative test is used as an aid in the diagnosis of an infectious disease, such as HIV, Streptococcus, or Trichomonas, the "cutoff" is the "limit of blank." (See the Definitions section and CLSI/NCCLS document EP17.<sup>5</sup>) The objective is to recognize the presence of the characteristic, such as immunologic response in the host or the presence of an antigen of the pathogen, which is indicative of