This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

November 2011

EP24-A2

Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition

This document provides a protocol for evaluating the accuracy of a test to discriminate between two subclasses of subjects when there is some clinically relevant reason to separate them. In addition to the use of receiver operating characteristic curves and the comparison of two curves, the document emphasizes the importance of defining the question, selecting the sample group, and determining the "true" clinical state.

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

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

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
standard@clsi.org

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

ISBN 1-56238-777-4 (Print)
ISBN 1-56238-778-2 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic)

EP24-A2
Vol. 31 No. 23
Replaces GP10-A
Vol. 15 No. 19

Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition

Volume 31 Number 23

Martin H. Kroll, MD
Bipasa Biswas
Jeffrey R. Budd, PhD
Paul Durham, MA
Robert T. Gorman, PhD
Thomas E. Gwise, PhD
Abdel-Baset Halim, PharmD, PhD, DABCC
Aristides T. Hatjimihail, MD, PhD
Jørgen Hilden, MD
Kyunghee Song, PhD

Abstract

Clinical and Laboratory Standards Institute document EP24-A2—*Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition* provides guidance for laboratorians and manufacturers who assess clinical test accuracy. It is not a recipe; rather, it is a set of concepts to be used to design an assessment of test performance or to interpret data generated by others. In addition to the use of ROC curves and comparison of two curves, the document emphasizes the importance of defining the question, selecting a sample group, and determining the "true" clinical state. The statistical data generated can be useful whether one is considering replacing an existing test, creating or adding a new test, or eliminating a current test.

Clinical and Laboratory Standards Institute (CLSI). *Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition.* CLSI document EP24-A2 (ISBN 1-56238-777-4 [Print]; ISBN 1-56238-778-2 [Electronic]). Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087 USA, 2011.

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. 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; Website: www.clsi.org.

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

Copyright ©2011 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.

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.

Suggested Citation

CLSI. *Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition.* CLSI document EP24-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2011.

Proposed Guideline
March 1987

Tentative Guideline
December 1993

Approved Guideline
December 1995

Approved Guideline—Second Edition
November 2011

ISBN 1-56238-777-4 (Print)
ISBN 1-56238-778-2 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic)

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

Committee Membership

Consensus Committee on Evaluation Protocols

Greg Cooper, CLS, MHA
Chairholder
W. Gregory Cooper LLC
Denton, Texas, USA

James F. Pierson-Perry
Vice-Chairholder
Siemens Healthcare Diagnostics
Newark, Delaware, USA

J. Rex Astles, PhD, FACB, DABCC
Centers for Disease Control and Prevention
Atlanta, Georgia, USA

Jeffrey R. Budd, PhD
Beckman Coulter
Chaska, Minnesota, USA

Jonathan Guy Middle, PhD
University Hospital Birmingham NHS Trust
Birmingham, United Kingdom

Mitchell G. Scott, PhD
Washington University School of Medicine
St. Louis, Missouri, USA

Lakshmi Vishnuvajjala, PhD
FDA Center for Devices and Radiological Health
Silver Spring, Maryland, USA

Document Development Committee on Assessment of Diagnostic Accuracy of Laboratory Tests Using ROC Curves

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

Jeffrey R. Budd, PhD
Beckman Coulter
Chaska, Minnesota, USA

Robert T. Gorman, PhD
Siemens Healthcare Diagnostics Inc.
Newark, Delaware, USA

Jørgen Hilden, MD
University of Copenhagen
Copenhagen, Denmark

Kyunghee Song, PhD
FDA Center for Devices and Radiological Health
Silver Spring, Maryland, USA

Staff

Clinical and Laboratory Standards Institute
Wayne, Pennsylvania, USA

Luann Ochs, MS
Vice President, Standards Development

Ron S. Quicho
Staff Liaison

Patrice E. Polgar
Project Manager

Megan P. Larrisey, MA
Editor

Acknowledgment

CLSI and the Consensus Committee on Evaluation Protocols gratefully acknowledge the following individuals for their help in preparing this document:

Bipasa Biswas
FDA Center for Devices and Radiological Health
Silver Spring, Maryland, USA

Paul Durham, MA
Culver City, California

Thomas E. Gwise, PhD
FDA Center for Drug Evaluation and Research
Silver Spring, Maryland, USA

Abdel-Baset Halim, PharmD, PhD, DABCC
Daiichi Sankyo Pharma Development
Edison, New Jersey, USA

Aristides T. Hatjimihail, MD, PhD
Hellenic Complex Systems Laboratory
Drama, Greece

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

Contents

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

Foreword

Laboratorians, investigators, *in vitro* diagnostic manufacturers, and clinicians are often interested in how well a test performs clinically. This is true whether considering replacing an existing test with a newer one, adding a new test to the laboratory's menu, eliminating tests where possible, or evaluating the diagnostic power of a laboratory test relative to another clinical or diagnostic tool. This project was originally intended to make recommendations about assessing the clinical performance of diagnostic tests. The concepts of Swets and Pickett[1] were adopted, whereby clinical performance is divided into (1) a discriminatory or diagnostic element (diagnostic accuracy) and (2) a decision or efficacy element. Laboratory tests are ordered to help answer questions about patient management. How much help an individual test result provides is variable and, in any case, a highly complicated issue. Management decisions and strategies are complex activities that require the physician to consider probabilities of disease, quality of the data available, effectiveness of various treatment/management alternatives, probability of outcomes, and value (and cost) of outcomes to the patient. Many types of clinical data (including laboratory results) are usually integrated into a complex decision-making process. Most often, a single laboratory test result is not the sole basis for a diagnosis or a patient-management decision.

Therefore, some have criticized the practice of evaluating the diagnostic performance of a test as if it were used alone. However, each clinical tool (eg, a clinical laboratory test, an electroencephalogram, an electrocardiogram, a nuclide scan, an X-ray, a biopsy, a pulmonary function test, or a sonogram) is meant to make some definable discrimination. It is important to know just how inherently accurate each test is as a diagnostic discriminator. *Note that assessing diagnostic accuracy, without engaging in comprehensive clinical decision analysis, is a valid and useful activity for the clinical laboratory.* Diagnostic accuracy is the most fundamental characteristic of the test itself as a classification device; it measures the ability of the test to discriminate among alternative states of health. In the simplest form, this property is the ability to distinguish between just two states of health or circumstances. Sometimes this involves distinguishing health from disease; other times it might involve distinguishing between benign and malignant disease, categorizing subjects as responding to therapy vs those not responding, or predicting who will become ill vs who will not. This ability to distinguish or discriminate between two states among subjects is a property of the test itself.

Indeed, the ability of the test to distinguish between the relevant alternative states or conditions of the subject (ie, diagnostic accuracy) is the most basic property of a laboratory test as a device to help in decision making. Note that this basic property cannot be separated from the clinical problem being addressed and the spectrum effect of the mix of subject states on which the test system is based. This property is the place to start when assessing the value of a test in the patient-management process.

Exploration of the usefulness of medical information, such as test data, involves a number of factors or parameters that are not properties of the test system; rather, they are properties of the circumstances of the clinical application. These include the probability or prevalence of disease, the possible clinical outcomes and the relative values of diagnostic outcomes, the costs to the patient (and others) of incorrect information (false-positive and false-negative classifications), and the costs and benefits of various treatment options. These characteristics or properties form the context in which test information is used, but are not properties of the test system. These factors interact with test results to affect the usefulness of the test, but do not affect test accuracy.

In summary, diagnostic accuracy is defined as the basic ability to discriminate between two subclasses of subjects when there is some clinically relevant reason to separate them. This concept of diagnostic accuracy refers to the quality of the information (classification) provided by the test, which should be distinguished from the practical usefulness of the information.[1] Both are aspects of test performance. The assessment of diagnostic accuracy is the place to start in evaluating test performance. If a test cannot discriminate between clinically relevant subclasses of subjects, then there is little incentive to further explore a possible clinical role. If, on the other hand, a test does exhibit a substantial ability to

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

discriminate, then by examining the degree of accuracy of the test and/or by comparing its accuracy to that of other tests, one can decide whether to delve into a more complex assessment of its role in patient management (decision analysis). This document addresses the assessment of diagnostic accuracy but not the analysis of usefulness or the role of the test in the patient-management process.

In this second edition of the guideline, the document development committee has provided more details on the construction and interpretation of receiver operating characteristic (ROC) curves. Many more examples are included to help the reader assess an individual curve and its associated area under the curve, as well as to compare two curves. Sample size calculations are provided for the first time.

NOTE: Although a step-by-step technique for generating ROC curves has been presented in EP24, it is assumed that most users of this guideline will access commercially available software for this task.

Key Words

Area under the curve, diagnostic accuracy, false-negative fraction, false-positive fraction, medical decision level, receiver operating characteristic curve, sensitivity, specificity, true-negative fraction, true-positive fraction

This is a preview of "CLSI EP24-A2". Click here to purchase the full version from the ANSI store.

Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition

1 Scope

This guideline outlines the steps and principles of prospectively planned and retrospective studies to evaluate the intrinsic diagnostic accuracy of a clinical laboratory test, defined as its fundamental ability to discriminate correctly among alternative states of health. It is not intended to help determine how best to use a diagnostic test in clinical practice, but instead to determine how accurate a laboratory test is in terms of diagnostic sensitivity and specificity.

Receiver operating characteristic (ROC) curve methodology arose in response to needs in electronic signal detection and problems with radar in the early 1950s.[2] It is derived from conditional probabilities, as originally formulated by Bayes.[3] This guideline aims to define ROC curves and to explain how to design, construct, interpret, and apply the information from ROC studies to evaluate diagnostic tests. For simplicity, only continuous scales, such as those typical for *in vitro* diagnostic tests, are discussed. The clinical condition that the test is intended to detect must be verifiable through some means other than the test under investigation. In other words, there must be an independent clinical reference standard against which one can compare the test. By selecting cutoffs between positive and negative diagnoses along the continuous scale of the test, the diagnostic outcomes for these decision levels are compared to the true clinical condition, which, in turn, generates the ROC curve.

This guideline will be of value to a wide variety of possible users, including:

- Investigators who are developing new tests for specific applications
- Manufacturers of reagents and devices for performing tests who are interested in assessing or validating test performance in terms of diagnostic accuracy
- Regulatory agencies interested in establishing requirements for claims related to diagnostic accuracy
- Clinical laboratorians who are reviewing data or the literature, and/or generating their own data, to make decisions about which tests to employ in their laboratories
- Health care or scientific workers interested in critical evaluation of data being presented on clinical test performance

2 Introduction

An ROC curve provides the following advantageous properties:

- It visually displays the performance of one or more diagnostic markers or tests across the entire measuring interval.
- By plotting unitless values (sensitivity vs specificity or sensitivity vs 1 − specificity), one can compare the diagnostic performance of two or more diagnostic markers or tests regardless of:
  – Units of expression of different markers or tools (eg, mg/dL, mmol/L, U/L)
  – Type of diagnostic test (eg, a clinical laboratory test, pulmonary function test, radiography)
  – Type of biological sample analyzed (eg, serum vs urine, saliva vs blood)

©*Clinical and Laboratory Standards Institute. All rights reserved.*