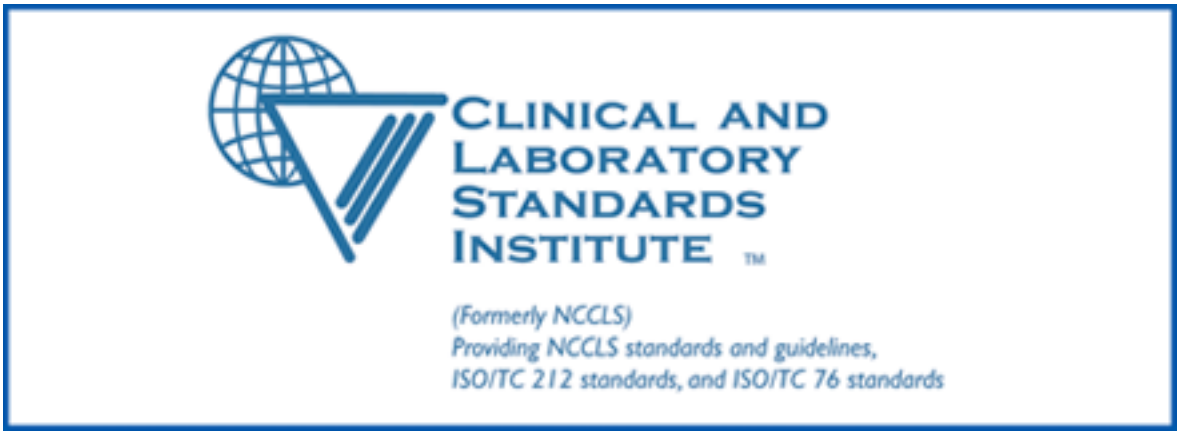
I/LA21-A
Vol. 22 No. 9

Replaces I/LA21-P
Vol. 19 No. 19

Clinical Evaluation of Immunoassays; Approved Guideline



This document addresses the need for clinical evaluation of new immunoassays and new applications of existing assays. As a guide to designing and executing a clinical evaluation, this document will aid clinical and regulatory personnel responsible for commercializing products, developers of "in-house" assays for institutional use, and developers of assays used for monitoring pharmacologic effects of new drugs or biologics.

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

NCCLS

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.

Clinical Evaluation of Immunoassays; Approved Guideline

Abstract

This document addresses the need for the clinical evaluation of new immunoassays and new applications of existing assays. Existing NCCLS documents provide guidance for assessing analytical performance, methods comparison, and clinical accuracy of laboratory tests. However, none of the documents define the elements that are integral to generating clinical data. As a guide to designing, executing, and analyzing a clinical evaluation, this document will aid clinical and regulatory personnel responsible for commercializing products, developers of "in-house" assays for institutional use, and developers of assays used for monitoring pharmacologic effects of new drugs or biologics.

The elements of this guideline include (1) a brief review of the analytical performance measures that must be in place prior to testing clinical specimens; (2) a thorough discussion of the planning and design considerations that are necessary for a successful evaluation; (3) a description of requirements for conducting the evaluation through monitoring and database management; and (4) a development plan for an effective analysis and evaluation.

NCCLS. *Clinical Evaluation of Immunoassays; Approved Guideline*. NCCLS document I/LA21-A (ISBN 1-56238-455-4). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.

I/LA21-A
ISBN 1-56238-455-4
ISSN 0273-3099

Clinical Evaluation of Immunoassays; Approved Guideline

Volume 22 Number 9

Linda Ivor, Chairholder
Peter Maxim, Ph.D.
Donald R. Parker, Ph.D.
Gregory P. Payne, RAC
John Ridderhof, Dr.P.H.
Carol Ryerson, Ph.D.
Lawrence E. Schaefer
Mark H. Zweig, M.D.

Suggested Citation

(NCCLS. *Clinical Evaluation of Immunoassays; Approved Guideline.* NCCLS document I/LA21-A [ISBN 1-56238-455-4]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.)

Proposed Guideline
September 1999

Approved Guideline
June 2002

ISBN 1-56238-455-4
ISSN 0273-3099

Committee Membership

Area Committee on Immunology and Ligand Assay

W. Harry Hannon, Ph.D. **Chairholder**	**Centers for Disease Control and Prevention** **Atlanta, Georgia**
Jeffrey Warren, M.D. **Vice-Chairholder**	**University of Michigan** **Ann Arbor, Michigan**

Subcommittee on Clinical Evaluations of Immunoassays

Linda Ivor **Chairholder**	**Gen-Probe Incorporated** **San Diego, California**
Peter Maxim, Ph.D.	FDA Center for Devices/Rad. Health Rockville, Maryland
Donald R. Parker, Ph.D.	Bayer Corporation Elkhart, Indiana
Gregory P. Payne, RAC	Hybritech San Diego, California
John C. Ridderhof, Dr.P.H.	Centers for Disease Control and Prevention Atlanta, Georgia
Carol Ryerson, Ph.D.	Eastman Kodak Co. Rochester, New York
Lawrence E. Schaefer	Abbott Laboratories Abbott Park, Illinois
Mark H. Zweig, M.D.	National Institutes of Health Bethesda, Maryland

Advisors

John A. Boffa, B.L.D.	American Association of Bioanalysts Howell, New Jersey
Prof. Aw Tar Choon	National University Hospital Singapore
Anna J. DeMarinis, M.T.(ASCP)	bioMerieux, Inc. Rockland, Massachusetts
Steven I. Gutman, M.D., MBA	FDA Center for Devices/Rad. Health Rockville, Maryland
Sheri L. Hall	Beckman Coulter, Inc. Brea, California

Advisors (Continued)

Lorin Hills, Ph.D.	LifeScan, Inc. Milpitas, California
Hassan Khayam-Bashi, Ph.D.	San Francisco General Hospital San Francisco, California
Edward M. Levine, Ph.D.	Diagnostic Products Corporation Los Angeles, California
Terry McGovern	Rockland, Massachusetts
Ronald W. McLawhon, M.D., Ph.D.	University of Chicago Medical Center Chicago, Illinois
William F. Skogen, Ph.D.	Dade Behring Inc. Glasgow Newark, Delaware
Leonard Wilson	Food and Drug Administration Rockville, Maryland
Patrice E. Polgar *Project Manager/Editor*	NCCLS Wayne, Pennsylvania
Donna M. Wilhelm *Assistant Editor*	NCCLS Wayne, Pennsylvania

Active Membership
(as of 1 April 2002)

Sustaining Members

Abbott Laboratories
American Association for Clinical Chemistry
Bayer Corporation
Beckman Coulter, Inc.
BD and Company
bioMérieux, Inc.
CLMA
College of American Pathologists
GlaxoSmithKline
Nippon Becton Dickinson Co., Ltd.
Ortho-Clinical Diagnostics, Inc.
Roche Diagnostics, Inc.

Professional Members

AISAR-Associazione Italiana per lo Studio degli
American Academy of Family Physicians
American Association for Clinical Chemistry
American Association for Respiratory Care
American Chemical Society
American Medical Technologists
American Public Health Association
American Society for Clinical Laboratory Science
American Society of Hematology
American Society for Microbiology
American Type Culture Collection, Inc.
Asociación Española Primera de Socorros (Uruguay)
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
CADIME-Camara De Instituciones De Diagnostico Medico
Canadian Society for Medical Laboratory Science—Société Canadienne de Science de Laboratoire Médical
Clinical Laboratory Management Association
COLA
College of American Pathologists
College of Medical Laboratory Technologists of Ontario
College of Physicians and Surgeons of Saskatchewan
ESCMID
Fundación Bioquímica Argentina
International Association of Medical Laboratory Technologists
International Council for Standardization in Haematology
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 Society for Histotechnology, Inc.
Ontario Medical Association Quality Management Program-Laboratory Service
RCPA Quality Assurance Programs PTY Limited
Sociedade Brasileira de Analises Clinicas
Sociedade Brasileira de Patologia Clinica
Sociedad Espanola de Bioquimica Clinica y Patologia Molecular
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
Health Care Financing Administration/CLIA Program
Health Care Financing Administration
Iowa State Hygienic Laboratory
Massachusetts Department of Public Health Laboratories
National Association of Testing Authorities – Australia
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
Ohio 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
Thailand Department of Medical Sciences

Industry Members

AB Biodisk
Abbott Laboratories
Abbott Laboratories, MediSense Products
Acrometrix Corporation
Advanced Medicine, Inc.
Agilent Technologies, Inc.
Ammirati Regulatory Consulting
Anaerobe Systems
Asséssor
AstraZeneca
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
BioMedia Laboratories Sdn Bhd
BioMérieux (NC)
bioMérieux, Inc. (MO)
Biometrology Consultants
Bio-Rad Laboratories, Inc.
Bio-Rad Laboratories, Inc. - France
Biotest AG
Blaine Healthcare Associates, Inc.
Bristol-Myers Squibb Company
Canadian External Quality Assessment Laboratory
Capital Management Consulting, Inc.
Carl Schaper
Checkpoint Development Inc.
Chiron Corporation
ChromaVision Medical Systems, Inc.
Chronolab Ag
Clinical Design Group Inc.
Clinical Laboratory Improvement Consultants
Cognigen
Community Medical Center (NJ)
Control Lab (Brazil)
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
Diagnostic Products Corporation
Eiken Chemical Company, Ltd.
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.
Helena Laboratories
Home Diagnostics, Inc.
Immunicon Corporation
Instrumentation Laboratory
International Technidyne Corporation
IntraBiotics Pharmaceuticals, Inc.
I-STAT Corporation
Johnson and Johnson Pharmaceutical Research and Development, L.L.C.
Kendall Sherwood-Davis & Geck
LAB-Interlink, Inc.
Laboratory Specialists, Inc.
Labtest Diagnostica S.A.
LifeScan, Inc. (a Johnson & Johnson Company)
Lilly Research Laboratories
Macemon Consultants
Medical Device Consultants, Inc.
Merck & Company, Inc.
Minigrip/Zip-Pak
Molecular Diagnostics, Inc.
mvi Sciences (MA)
Nabi
Nichols Institute Diagnostics (Div. of Quest Diagnostics, Inc.)
Nissui Pharmaceutical Co., Ltd.
Nippon Becton Dickinson Co., Ltd.
Norfolk Associates, Inc.
Novartis Pharmaceuticals Corporation
Ortho-Clinical Diagnostics, Inc. (Raritan, NJ)
Ortho-Clinical Diagnostics, Inc. (Rochester, NY)
Oxoid Inc.
Paratek Pharmaceuticals
Pfizer Inc
Pharmacia Corporation
Powers Consulting Services
Premier Inc.
Procter & Gamble Pharmaceuticals, Inc.
The Product Development Group
QSE Consulting
Quintiles, Inc.
Radiometer America, Inc.
Radiometer Medical A/S
David G. Rhoads Associates, Inc.
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
Showa Yakuhin Kako Company, Ltd.
Streck Laboratories, Inc.
SurroMed, Inc.
Sysmex Corporation (Japan)
Sysmex Corporation (Long Grove, IL)
The Clinical Microbiology Institute
The Toledo Hospital (OH)
Trek Diagnostic Systems, Inc.
Versicor, Inc.
Vetoquinol S.A.
Visible Genetics, Inc.
Vysis, Inc.
Wallac Oy
Wyeth-Ayerst
Xyletech Systems, Inc.
YD Consultant
YD Diagnostics (Seoul, Korea)

Trade Associations

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

Associate Active Members

20th Medical Group (SC)
31st Medical Group/SGSL (APO, AE)
67th CSH Wuerzburg, GE (NY)
121st General Hospital (CA)
Academisch Ziekenhuis-VUB (Belgium)
Acadiana Medical Laboratories, LTD (LA)
Adena Regional Medical Center (OH)
Advocate Healthcare Lutheran General (IL)
Akershus Central Hospital and AFA (Norway)
Albemarle Hospital (NC)
Allegheny General Hospital (PA)

Allegheny University of the Health Sciences (PA)
Allina Health System (MN)
Alton Ochsner Medical Foundation (LA)
American Medical Laboratories (VA)
Arkansas Department of Health
ARUP at University Hospital (UT)
Armed Forces Research Institute of Medical Science (APO, AP)
Associated Regional & University Pathologists (UT)
Aurora Consolidated Laboratories (WI)
Azienda Ospedale Di Lecco (Italy)
Bay Medical Center (MI)
Baystate Medical Center (MA)
Bbaguas Duzen Laboratories (Turkey)
Bermuda Hospitals Board
Bo Ali Hospital (Iran)
British Columbia Cancer Agency (Vancouver, BC, Canada)
Broward General Medical Center (FL)
Calgary Laboratory Services
Carilion Consolidated Laboratory (VA)
Cathay General Hospital (Taiwan)
CB Healthcare Complex (Sydney, NS, Canada)
Central Peninsula General Hospital (AK)
Central Texas Veterans Health Care System
Centre Hospitalier Regional del la Citadelle (Belgium)
Centro Diagnostico Italiano (Milano, Italy)
Champlain Valley Physicians Hospital (NY)
Changi General Hospital (Singapore)
Children's Hospital (NE)
Children's Hospital & Clinics (MN)
Children's Hospital King's Daughters (VA)
Children's Hospital Medical Center (Akron, OH)
Children's Hospital of Philadelphia (PA)
Children's Medical Center of Dallas (TX)
Clarian Health–Methodist Hospital (IN)
Clendo Lab (Puerto Rico)
Clinical Laboratory Partners, LLC (CT)
CLSI Laboratories (PA)
Columbia Regional Hospital (MO)
Commonwealth of Kentucky
Community Hospital of Lancaster (PA)
CompuNet Clinical Laboratories (OH)
Cook Children's Medical Center (TX)
Covance Central Laboratory Services (IN)
Danish Veterinary Laboratory (Denmark)
Danville Regional Medical Center (VA)
Delaware Public Health Laboratory
Department of Health & Community Services (New Brunswick, Canada)
DesPeres Hospital (MO)
DeTar Hospital (TX)
Detroit Health Department (MI)
Dr. Everett Chalmers Hospital (New Brunswick, Canada)
Doctors Hospital (Bahamas)
Duke University Medical Center (NC)
E.A. Conway Medical Center (LA)
Eastern Maine Medical Center
East Side Clinical Laboratory (RI)
Elyria Memorial Hospital (OH)
Emory University Hospital (GA)
Esoterix Center for Infectious Disease (TX)
Fairview-University Medical Center (MN)
Federal Medical Center (MN)
Florida Hospital East Orlando
Foothills Hospital (Calgary, AB, Canada)
Fort St. John General Hospital (Fort St. John, BC, Canada)
Fox Chase Cancer Center (PA)
Fresenius Medical Care/Spectra East (NJ)
Fresno Community Hospital and Medical Center
Frye Regional Medical Center (NC)
Gambro Healthcare Laboratory Services (FL)
Gateway Medical Center (TN)
Geisinger Medical Center (PA)
Grady Memorial Hospital (GA)
Guthrie Clinic Laboratories (PA)
Hahnemann University Hospital (PA)
Harris Methodist Erath County (TX)
Harris Methodist Fort Worth (TX)
Hartford Hospital (CT)
Headwaters Health Authority (Alberta, Canada)
Health Network Lab (PA)
Health Partners Laboratories (VA)
Heartland Regional Medical Center (MO)
Highlands Regional Medical Center (FL)
Hoag Memorial Hospital Presbyterian (CA)
Holmes Regional Medical Center (FL)
Holzer Medical Center (OH)
Hospital for Sick Children (Toronto, ON, Canada)
Hospital Sousa Martins (Portugal)
Hotel Dieu Hospital (Windsor, ON, Canada)
Houston Medical Center (GA)
Huddinge University Hospital (Sweden)
Hurley Medical Center (MI)
Indiana State Board of Health
Indiana University
Institute of Medical and Veterinary Science (Australia)
Instituto Scientifico HS. Raffaele (Italy)
International Health Management Associates, Inc. (IL)
Jackson Memorial Hospital (FL)
Jersey Shore Medical Center (NJ)
John C. Lincoln Hospital (AZ)
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)
Keller Army Community Hospital (NY)
Kenora-Rainy River Regional Laboratory Program (Ontario, Canada)
Kern Medical Center (CA)
Kimball Medical Center (NJ)
King Faisal Specialist Hospital (Saudi Arabia)
King Khalid National Guard Hospital (Saudi Arabia)
King's Daughter Medical Center (KY)

Klinični Center (Slovenia)
Laboratories at Bonfils (CO)
Laboratoire de Santé Publique du Quebec (Canada)
Laboratório Fleury S/C Ltda. (Brazil)
Laboratory Corporation of America (NJ)
Laboratory Corporation of America (MO)
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)
Libero Instituto Univ. Campus BioMedico (Italy)
Long Beach Memorial Medical Center (CA)
Louisiana State University Medical Center
Maccabi Medical Care and Health Fund (Israel)
Magee Womens Hospital (PA)
Malcolm Grow USAF Medical Center (MD)
Manitoba Health (Winnipeg, Canada)
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
Mescalero Indian Hospital (NM)
Methodist Hospital (TX)
Methodist Hospitals of Memphis (TN)
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)
MRL Pharmaceutical Services, Inc. (VA)
MRL Reference Laboratory (CA)
Nassau County Medical Center (NY)
National Institutes of Health (MD)
Naval Surface Warfare Center (IN)
Nebraska Health System
New Britain General Hospital (CT)
New England Fertility Institute (CT)
North Carolina State Laboratory of Public Health
Northern Indiana Education Foundation
North Kansas City Hospital (MO)
North Shore – Long Island Jewish Health System Laboratories (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)
Our Lady of Lourdes Hospital (NJ)
Our Lady of the Resurrection Medical Center (IL)
Pathology and Cytology Laboratories, Inc. (KY)
The Permanente Medical Group (CA)
Piedmont Hospital (GA)
Pikeville Methodist Hospital (KY)
Pocono Hospital (PA)
Presbyterian Hospital of Dallas (TX)
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
Riyadh Armed Forces Hospital (Saudi Arabia)
Royal Columbian Hospital (New Westminster, BC, Canada)
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. Joseph Hospital (NE)
St. Joseph's Hospital – Marshfield Clinic (WI)
St. Joseph Mercy Hospital (MI)
St. Luke's Regional Medical Center (IA)
St. Mary Medical Center (IN)
St. Mary of the Plains Hospital (TX)
St. Mary's Hospital & Medical Center (CO)
St. Paul's Hospital (Vancouver, BC, Montreal)
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)
Seoul Nat'l University Hospital (Korea)
Shanghai Center for the Clinical Laboratory (China)
South Bend Medical Foundation (IN)
Southwest Texas Methodist Hospital (TX)
South Western Area Pathology Service (Australia)
Southern Maine Medical Center
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)

Sunrise Hospital and Medical Center (NV)
Swedish Medical Center – Providence Campus (WA)
Tampa General Hospital (FL)
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)
UCSF Medical Center (CA)
UNC Hospitals (NC)
University Hospital (Gent) (Belgium)
University Hospitals of Cleveland (OH)
The University Hospitals (OK)
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 Texas M.D. Anderson Cancer Center
University of Virginia Medical Center
University of Washington
UZ-KUL Medical Center (Belgium)
VA (Denver) Medical Center (CO)
VA (Kansas City) Medical Center (MO)
VA (Western NY) Healthcare System
VA (San Diego) Medical Center (CA)
VA (Tuskegee) Medical Center (AL)
VA Outpatient Clinic (OH)
Vejle Hospital (Denmark)
Washington Adventist Hospital (MD)
Washoe Medical Center Laboratory (NV)
West Jefferson Medical Center (LA)
Wilford Hall Medical Center (TX)
William Beaumont Hospital (MI)
Williamsburg Community Hospital (VA)
Winn Army Community Hospital (GA)
Winnipeg Regional Health Authority (Winnipeg, Canada)
Wishard Memorial Hospital (IN)
Yonsei University College of Medicine (Korea)
York Hospital (PA)

Contents

Foreword

Currently, no uniform guidelines exist that adequately describe the requirements for the clinical evaluation of immunoassays. Historically, assay developers—primarily *in vitro* diagnostic manufacturers—have based their approach to designing and conducting clinical evaluations on what may be required by government regulatory agencies, who review the demonstration of safety and effectiveness. Increasingly, assays developed by nongovernment-regulated entities are being used as clinical measures. These assays may include those developed for the purpose of measuring end points of drugs under development. While these end-point assays may provide appropriate analysis of analytical performance that is consistent with laboratory requirements and regulations, evaluation of clinical performance may be incomplete or lacking.

In preparing this guideline, the subcommittee considered three areas of need regarding the clinical use of immunoassays. First, for manufacturers of *in vitro* diagnostic assays, this guideline will provide a checklist to review against their approach to addressing regulatory requirements for commercialization of products. Second, for laboratories engaged in the development of immunoassays for use within their institutions, this guideline will provide direction in designing an evaluation of the assay's clinical performance. And third, for those scientists involved in evaluating new therapeutic agents, this guideline will provide direction in establishing immunoassays as reliable clinical end points.

For the purposes of this document, clinical performance refers to accuracy— correct classification, i.e., clinical sensitivity and specificity—and does not refer to clinical utility, which may include the effects of environment, economy, and patient outcomes. While there is mention of an assay's analytical performance, users should refer to existing NCCLS documents (see Related NCCLS Publications) and to other sources for more detailed information.

Because the scope of this document does not limit its application to industry or to the clinical or research laboratory, the subcommittee has used the term "clinical evaluation" in place of "clinical study" or "clinical trial." While considered interchangeable from the subcommittee's perspective, the reader should use the term that is appropriate for the reader's institution.

It should also be acknowledged that there are different types of evaluations for new assays, including comparative and clinical. Comparative evaluations are typically performed when the laboratorian is considering substituting an assay from one manufacturer with another from a different manufacturer. While having its own unique forms of execution and analyses, this evaluation is simply comparing one assay to another without the postulation of any clinical questions. A reference for comparative evaluations is NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples*. While it may involve a comparative approach, the clinical evaluation is required for the application of a new assay, a new analyte, or for a new, intended use of an existing analyte.

In the assay development to implementation/commercialization continuum, this guideline addresses the activities associated with preclinical testing and clinical evaluation requirements, evaluation design, and analysis. While written for immunoassay developers, the information has broad application to other assay formats. It is the subcommittee's intent that this document will be expanded to ensure that the full range of *in vitro* diagnostic assays is addressed.

A Note on Terminology

NCCLS, as a global leader in standardization and harmonization, is firmly 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 must be an evolutionary and educational process that begins with new projects and revisions of existing documents.

In the context of this document, it is necessary to point out that several terms are used differently in the USA and other countries, notably those in Europe.

In Europe the term "performance evaluation" is used for the assessment of quality of *in vitro* diagnostic products both with their analytical and medical (diagnostic) characteristics. "Clinical evaluation" in European terms is applied mostly to the evaluation of medical products, which are used on or in patients or when it refers to clinical studies of drugs, under much more stringent conditions. Appropriately, the USA term "clinical evaluation" in the context of this document would translate into "diagnostic evaluation" in Europe. Consequently, the terms "diagnostic sensitivity" and "diagnostic specificity" are used in Europe, with the corresponding expressions "clinical sensitivity" and "clinical specificity" in the USA, as they are applied in this document.

Also, in order to align the usage of terms to ISO, the term "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. The term "accuracy," in its metrological sense, refers to the closeness of the agreement between the result of a (single) measurement and a true value of a measurand, thus comprising both random and systematic effects.

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 NCCLS document M29—*Protection of Laboratory Workers from Occupationally Acquired Infections*.

Key Words

Clinical evaluation, diagnostic evaluation, clinical evaluation investigator, clinical performance characteristics, diagnostic performance characteristics, database management, evaluation population, informed consent, institutional review board, pilot evaluation, sponsor, statistical tests

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:

QSEs

Documents & Records	Information Management
Organization	Occurrence Management
Personnel	Assessment
Equipment	Process Improvement
Purchasing & Inventory	Service & Satisfaction
Process Control	Facilities & Safety

NCCLS document I/LA21-A—*Clinical Evaluation of Immunoassays* addresses the following Quality System Essentials (QSEs)

Documents & Records	Organization	Personnel	Equipment	Purchasing & Inventory	Process Control	Information Management	Occurrence Management	Assessment	Process Improvement	Service & Satisfaction	Facilities & Safety
X			X		X	X					

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

Clinical Evaluation of Immunoassays; Approved Guideline

1 Introduction

In vitro diagnostic assay development occurs in multiple environments, each with different requirements for the actual use of the assay. Developers that manufacture and market their assays are required to obtain governmental approval or clearance in some countries, such as in the U.S., prior to commercial sales. Others may develop assays that do not require the same extent of testing and review. For example, clinical research laboratories may develop assays for use within their institutions, thereby limiting the assay's use to a select patient population. Another example is the development of research assays for use in monitoring pharmacologic effects or unfavorable reactions. Whatever the intended use, the assay can only provide medical benefit when it has undergone adequate clinical evaluation.

This document provides uniform guidelines for clinical evaluation of *in vitro* diagnostic immunoassays. Separated into four basic sections, this guidance is intended to provide all assay developers with a consistent approach to establishing clinical performance characteristics. First, the preclinical evaluation is described, including recommendations for establishing analytical performance. Second, the planning and design of the clinical evaluation is discussed. Third, a description of conducting the evaluation is given, and fourth, analysis and summary documentation is reviewed.

1.1 Scope

The scope of this document is to provide guidance for evaluating the clinical performance characteristics of an immunoassay, with no regard to the status or environment of the assay developer. The information provided will have broad applications to multiple assay formats and their uses.

1.2 Definitions[a]

Accuracy// Measurement accuracy// Accuracy of measurement, *n* - Closeness of the agreement between the result of a measurement and a true value of the measurand {/analyte}; See **Trueness**.

Analytical specificity, *n* - The ability of a measurement procedure to determine solely the measurable quantity it purports to measure.

Clinical evaluation, *n* - An investigation of the clinical performance characteristics of a new (or new indication for use) *in vitro* diagnostic assay in controlled clinical settings; NOTE**:** The term "clinical evaluation" is equivalent to the term "diagnostic evaluation."

Clinical feasibility/Pilot evaluation, *n* – An evaluation performed using patient specimens to assess the potential application of a new assay to some clinical use; **NOTE:** Typically conducted by the sponsor, the evaluation may take place in a clinical setting or in the sponsor's laboratory.

Clinical investigator, *n* – A person under whose direction a clinical evaluation is conducted.

Clinical sensitivity, *n* - The proportion of patients with a well-defined clinical disorder whose test values are positive or exceed a defined decision limit (i.e., a positive result and identification of the patients who have a disease); **NOTES:** a) It is the fraction of clinically true positive classifications divided by the sum

[a] Some of these definitions are found in NCCLS document NRSCL8—*Terminology and Definitions for Use in NCCLS Documents.* For complete definitions and detailed source information, please refer to the most current edition of that document.

of clinically true positive plus clinically false negative classifications; b) The term "Clinical sensitivity" is equivalent to "Diagnostic sensitivity."

Clinical specificity, *n* - The proportion of subjects who do not have a specified clinical disorder and whose test results are negative or within the defined decision limit; **NOTES:** a) It is the fraction of clinically true negative classifications divided by the sum of clinically true negative plus clinically false positive classifications; b) The term "clinical specificity" is equivalent to "diagnostic specificity."

Confidence interval, *n* **-** An interval estimate of a population parameter computed so that the statement "the population parameter lies in this interval" will be true ... in a stated proportion of the times such statements are made.

Interim analysis, *n* - Any data analysis that is performed during the clinical evaluation and before the evaluation is completed; **NOTES:** a) Analyses performed at intervals throughout the evaluation provide the sponsor with estimates of the new assay's performance; b) These estimates are of limited validity due to insufficient sample size.

Minimal detectable concentration (MDC), *n* – The lowest concentration that can be accurately estimated by a given test system; **NOTE:** Sometimes this is defined as the lowest concentration that can be statistically distinguished from zero, but other definitions may be used.

Negative predictive value, NPV, *n* – The likelihood that an individual with a negative test result does not have the disease, or other characteristic which the test is designed to detect. **NOTE**: This varies with prevalence of the disease unless the test is 100% sensitive.

Population//(study group), *n* - The totality of items under consideration; **NOTE:** The study group is the sample subset of the population which is actually studied.

Positive predictive value, PPV, *n* – The likelihood that an individual with a positive test result has a particular disease, or characteristic, which the test is designed to detect; **NOTE:** This varies with prevalence of the disease unless the test is 100% specific.

Preclinical evaluation, *n* - A preliminary evaluation to determine if the new assay's analytical performance characteristics meet the desired design goals; **NOTE**: It serves as a determinant for proceeding to the clinical evaluation.

Precision, *n* – The closeness of agreement between independent test results obtained under prescribed {/stipulated} conditions; **NOTE:** Precision of a given measurement procedure is usually subdivided according to specific conditions into **Repeatability** and **Reproducibility**. Please refer to these terms below.

Repeatability of measurements {/Measurement repeatability}, *n* - The closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement; **NOTE:** Repeatability is often termed in laboratory medicine as "within-run precision."

Reportable range, *n* - The range of test values over which the relationship between the instrument, kit, or system's measurement response is shown to be valid; **NOTES:** a) For this document, the range of values (in units appropriate for the analyte) 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) This is similar to the VIM definition for "measuring range" or "working range," i.e., a set of values of measurands for which the error of a measuring instrument is intended to lie within specified limits; c) The reportable range of the assay should be established prior to beginning the clinical evaluation.

Reproducibility, of measurements {/Measurement reproducibility}, *n* **-** The closeness of the agreement between the results of measurements of the same measurand, where the measurements are carried out under changed conditions; **NOTE:** The changed conditions may refer different lots, runs, time (day), technician, etc.

Subject, *n* - A person, with disease or without disease, enrolled in an evaluation; **NOTE:** An equivalent term is "Proband."

Traceability, *n* - A property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons, all having stated uncertainties; **NOTES:** a) The concept is often expressed by the adjective "traceable"; b) The unbroken chain of comparisons is called a traceability chain**.**

Trueness/Trueness of measurement, *n* - Closeness of agreement between the average value obtained from a large series of results of measurements and a true value; **NOTE**: The difference between the average value from a large series of measurements and the true is also called "bias" or systematic error. See **Accuracy**.

2 Establishment of Analytical Performance Prior to Clinical Evaluations

Although the focus of this document is on the clinical evaluation of immunoassays, the information provided will have broad applications to multiple assay formats and their uses. It is imperative that the assay be well designed and optimized prior to starting a clinical evaluation. Therefore, before beginning a clinical evaluation, the assay should be produced as pilot lots according to prepared documentation, and not in a research and development modality. Modifications of the product design, materials, or operating conditions during the course of a clinical evaluation may invalidate the entire evaluation, due to the fact that the analytical performance characteristics of the assay may have changed while the evaluation was in progress. Such a change can significantly affect the variability of test results and may result in an inability to validate clinical performance. The key elements in the analytical characterization of an assay are briefly described below with references provided for more detailed information.

2.1 Assay Components

Immunoassays refer to any laboratory method for detecting a substance, or antigen, by using an antibody that is reactive with it. It is important to recognize that some immunoassays claiming to measure the same analyte may in fact give different results when applied to a particular sample or reference material because respective reagents recognize a different repertoire of epitopes in the substance thus leading to results for different although related quantities.[1] In some cases the substances are mixtures of molecular species with clinically relevant properties in common, but with different structures in varying proportions, e.g., substances involving analytes such as antibodies toward antigens, glycoproteins, tumor antigens, and protein hormones.[1] For many analytes measured by immunoassays, there is no traceability for trueness above the manufacturer's selected measurement procedure or working calibrator until internationally agreed upon reference measurements and materials become available.[1] The following section addresses the characterization of antigen and antibody, and calibration, as well as preparation and handling of other assay components.

2.1.1 Antigen and Antibody Characterization

Whether the antigen and antibody are produced internally or purchased from a vendor, it is important to ensure that an adequate supply of these critical reagents is available to avoid frequent material changes to new lots. Antigen characterization using biochemical, immunochemical, or immunological techniques is necessary to demonstrate purity, the nature and possibly quantity of potentially cross-reacting antigens, and the presence of all essential reactants in the antigen preparation. For example, it may be important

that proteins of 34, 69, and 157 kD all be present in the antigen. If the antigen is prepared from cell culture or by recombinant procedures, appropriate documentation should be available. Antibody should be tested by the most sensitive procedure available to determine the specificity and to characterize cross-reactivity with related antigens. A popular technique for establishing antibody specificity is western immunoblotting using a crude antigen mixture. In some cases, additional information, such as calculation of binding constants, etc., may be needed. If monoclonal antibodies are used, there should be adequate documentation of the procedures used for production of the hybridoma, as well as selection and expansion of the monoclonal antibody. If these reagents are purchased, sufficient characterization may have already been done, and the needed documentation can be supplied by the vendor. To use these reagents, the assay developer must establish acceptance criteria that will be used to release each new lot of antigen or antibody.

2.1.2 Calibrators, Standards, and Controls

Calibrators are materials with established analyte values that are used to standardize the instrument or assay method. Prepared at specific concentration intervals, the calibrators establish the assay dose response curve and should encompass the critical determination points. For example, if the key medical decision point for the assay is 2.0 ng/mL, there should be a calibrator close to that value.

Assay controls are materials that contain assay-specific analyte in an assay-compatible matrix and are tested with patient samples to ascertain the reliability of the assay. Rather than a preset value, controls may have a range of values and variability that should be consistent with the required accuracy for the assay working range. Generally, positive and negative controls are used for qualitative assays, and low- and high-level controls are used for the quantitative assays. In quantitative assays, there should be one control with analyte concentration near the decision point. A narrative description of these materials can be found in the most current edition of NCCLS document I/LA18—S*pecifications for Immunological Testing for Infectious Diseases.*

In addition to assay controls provided by the assay manufacturer or by the laboratory, commercial controls are available for assay developers and assay users. Typically, these controls contain different concentrations of one or several different analytes and may be used in addition to the assay controls to validate each assay run. In addition to use in assay validation, commercial controls may be used by the assay developer to challenge laboratory testing personnel's skills in the performance of an assay. These challenge controls are not used to aid directly in clinical diagnosis. Commercial controls are assigned specified control ranges that may be determined by the manufacturer of the testing methodology. These ranges are specified in the commercial controls package insert for each level of control. Availability of commercial controls during the assay development process provides the developer with a measure of the product's trueness and precision. If using unassayed controls, developers should establish acceptability limits prior to beginning the evaluation.[b]

National or international standards or reference materials provide the basis for calibration and may be used by assay developers and research or clinical laboratories to ensure harmonization across assays. When such suitable materials are available, a developer should use these materials to set values (establish traceability) for the assay calibrators, and an assay user can confirm that the product is in fact standardized to a particular reference material. In some cases, reference materials and procedures may not exist. For these assays, the developer must establish a means of ensuring purity of the calibration material and its preservation in a stable matrix, as well as a means of establishing its concentration.

[b] Refer to FDA guideline, "Points to Consider Guidance Document on Assayed and Unassayed Quality Control Material." Draft February 1999.

The trueness of the measurement of a value to a defined calibrator depends upon the metrological traceability of the value.[1] When possible, the calibration should be traced to a method and/or material of a higher metrological order/hierarchy.[1] The goals of calibration hierarchies are to improve traceability of results. The hierarchy of traceability depends upon on the availability of components for each of the various metrological levels of measurement procedures and calibrators; five levels of this traceability chain can be identified.[1] The metrological traceability chain ranges from the first level of availability of primary reference methods and primary reference materials for value assignment through several different iterations of these in combinations to the last level where no reference sources are available.[1] When neither reference measurement procedure nor reference materials for calibration are available, the manufacturer can establish "in-house" measurement procedures and calibrators to support their value assignment to their product's calibrator. This condition may occur frequently for immunoassays and traceability may be particularly difficult with these immunoprocedures.

2.1.3 Other Reagents

Reagents that will be provided with the assay, such as buffers, enzyme substrates, and stopping reagents, will need to be provided in the most practical and stable form (liquid or liquid concentrate, lyophilized, powder). Specifications will need to be established for those materials that are needed to run the assay but will not be provided with the assay (i.e., microplates, spectrophotometers, and water baths).

2.1.4 Stability, Handling, and Storage

Generally, real-time stability studies are performed to establish the expiration dating for the entire kit and its individual components under recommended storage conditions, and this information is provided on the label. Alternatively, accelerated stability studies can be conducted at elevated temperatures for materials that are suitable for testing by these methods. Assay components with unique storage requirements, such as reduced shelf-life after reconstitution, aliquoting and freezing after reconstitution, or storage in reduced light, should have labeling that indicates any special handling.

Warnings and precautions may be required for some reagents. Materials containing metallic azides have to be identified and proper disposal conditions described. Special warnings are required for any kit that contains radioactive materials. Kits that contain reagents made with human source materials must contain a statement that the source materials were tested and found negative for the presence of HIV and hepatitis antigens. Additional warnings and handling recommendations may be necessary depending on the kit contents. Material Safety Data Sheets should be available for hazardous materials.

2.2 Specimen Requirements

2.2.1 Type

Serum is the most commonly recommended specimen for immunoassay. It should be clearly stated if hemolysis, lipemia, cloudiness, or any other sample conditions are to be avoided. When plasma is included as a specimen, the recommended anticoagulants should also be listed, as some may interfere in assay performance. Comparison studies with paired samples should be performed to show that the same results are obtained with each sample type. Similar studies are needed to show that whole blood or urine or another specimen type is acceptable. Any special caveats regarding sample collection should be provided also. NCCLS document H3— *Procedures for the Collection of Diagnostic Blood Specimens by Venipuncture* offers additional details.

2.2.2 Stability, Handling, and Storage

It is incumbent on the assay developer or manufacturer to establish the conditions and time frame whereby a sample or specimen can be stored. These data to support stability claims can be generated

from real-time studies or might come from the literature or historical references, e.g. Tietz.[2] All patient specimens are to be treated with standard precautions, because it is impossible to know whether they are infectious. NCCLS document M29— *Protection of Laboratory Workers from Occupationally Acquired Infections* deals specifically with this issue.

2.3 Procedural Design

2.3.1 Assay Method Optimization

Optimization of the immunoassay can be considered to encompass all aspects of the test system. First, reproducibility should be established for the reaction conditions, linearity, working range, analytical specificity, minimal detectable concentration, etc. These operating conditions are then incorporated into a well-defined, step-by-step procedure. Second, the reference interval (normal range) of analyte levels, the expected values in appropriate test populations for which the assay is intended to be used, the key medical decision points, and/or cutoff are estimated. These values will be validated and challenged in the clinical studies. NCCLS document EP10— *Preliminary Evaluation of Quantitative Clinical Laboratory Methods* gives additional guidance on experimental design and data analysis for the preliminary evaluation of analytical device performance.

2.3.2 QA/QC

Ensuring that appropriate quality assurance and quality control procedures are followed requires the availability of the materials and protocols necessary for proper and accurate test performance. These include control materials and concentrations for positive and negative controls; any additional control parameters that need to be addressed; established QA criteria; directions for performing QC; interpretation of QC; and recommendations for action regarding discrepant QC results. If the device contains internal or procedural controls, these should be described with an explanation of the assay performance parameters that they control.

2.3.3 Training

The magnitude of training and skill level required to perform the assay should be delineated. Depending on the complexity of the test, training can range from a detailed, step-by-step procedure that is provided as a part of the product labeling, to an in-depth user manual, or to a formal on-site training session with practice specimens and certification requirements.

2.3.4 Ancillary Equipment

If the assay is not being developed as a part of an instrument system, the specifications for ancillary equipment, such as washers, diluters, readers, data management, and processing hardware and software, must be provided.

2.4 Analytical Measures

2.4.1 Performance Characteristics

The nonclinical laboratory testing is typically conducted prior to initiating a clinical study in order to establish the performance characteristics of the new assay. The specific experiments performed may vary, depending on whether the assay is quantitative, semiquantitative, or qualitative. Some of these studies are described in detail in other NCCLS evaluation protocols: EP5—*Evaluation of Precision Performance of Clinical Chemistry Devices*; EP6—*Evaluation of the Linearity of Quantitative Analytical Methods*; EP9—*Method Comparison and Bias Estimation Using Patient Samples;* and EP14—*Evaluation of Matrix Effects*.

The following is a general list of studies which have been used to establish immunoassay performance characteristics:

- validation of cutoffs and/or calibration curve for a qualitative assay;
- determination of lower and upper limits of detection (dynamic range) and linearity for a quantitative or semiquantitative assay; and
- precision, anticoagulant study, cross-reactivity, interference testing, minimal detectable concentration, and bias relative to a comparator method (or traceability to a reference measurement procedure[1]).

It is important to conduct evaluations of trueness and precision (repeatability and reproducibility) at each clinical evaluation site to ensure the comparability of clinical test data prior to testing clinical specimens.

2.4.2 Interference Studies

The common interfering substances tested during assay development include hemoglobin, lipids, bilirubin, and protein concentration. Other substances that have been investigated include therapeutic drugs, over-the-counter medications, and dietary supplements. Additional potential interferences will be highly dependent on the assay being developed, the specimen type, and the proposed test population. For additional guidance, the reader is referred to NCCLS documents EP7—*Interference Testing in Clinical Chemistry* and I/LA18—*Specifications for Immunological Testing for Infectious Diseases.*

2.4.3 Cross-Reactivity Studies

Cross-reactivity studies should be conducted to define the analytical specificity of the immunoassay for the specific antibody or antigen the assay is designed to detect. The experiment is conducted with a large excess of a variety of natural substances and chemical analogs to challenge the assay's specificity. The test substances examined vary greatly depending on the assay design. Substances that have been investigated include patient sera from subjects with related viruses or diseases, microorganisms that could be encountered in test specimens, and microorganisms which could be introduced when handling test specimens. If the antigen or antisera utilized in the assay is a recombinant protein, consider testing sera containing antibodies against the organism in which the vectors were induced.

3 Clinical Evaluation: Planning and Design

Adequate organization and communication are key elements to any clinical evaluation. Initial face-to-face interaction between the clinical evaluation sponsor or assay developer and the clinical investigator prior to beginning the evaluation establishes a productive working relationship which will prove invaluable throughout the entire clinical evaluation period. As a part of this initial interaction, the clinical investigator should be provided with the appropriate materials necessary to plan for, implement, and complete a proper clinical evaluation. The evaluation sponsor has the responsibility for organizing and providing such materials in an "investigator's manual." In addition, the sponsor must be familiar with and must fully inform the investigator of the appropriate ethical considerations relative to the evaluation being undertaken.

3.1 Investigator's Manual

It is recommended that the evaluation sponsor prepare an investigator's manual and provide it to each evaluation site. It should provide a set of clear and concise instructions which will serve as a reference guide during the clinical evaluation period. This manual may be used in conjunction with frequent person-to-person communications between the sponsor and investigator as the evaluation proceeds. It is

recommended that the manual be titled with the name of the immunoassay being evaluated, the names of the site and investigator, and the name and phone/fax numbers of the sponsor. Optional sponsor names and numbers may be provided for evening and weekend emergencies. The manual shall include, at a minimum, the following sections:

(1) **Clinical protocol**. A dated and revision-number-controlled copy of the clinical evaluation protocol. (See Section 3.3.)

(2) **Draft package insert**. A draft-written description of the immunoassay performance instructions, including equipment required, specimen collection/handling, and precautions, as well as a technical description of the new assay method and its principles of performance.

(3) **Institutional review board (IRB)**. Where appropriate, document that IRB review has occurred, along with a copy of an approved, blank, patient informed consent form.

(4) **Data collection**. Instructions describing the manner in which data (including patient histories, if appropriate) are to be collected, documented, and transmitted to the sponsor.

(5) **Forms management**. Copies of forms to be used during the evaluation, with examples of each properly filled out.

(6) **Daily work organization**. Suggestions for organization of daily work load related to the clinical evaluation (e.g., equipment maintenance required daily/weekly, recommendations for appropriate number of patient specimen runs per day, daily paperwork completion schedule, and preparation schedule for data/materials shipments back to the sponsor).

(7) **Supply inventory control**. Suggestions for maintaining adequate levels of supplies required for the evaluation, including instructions for ordering additional supplies/materials from either the sponsor or outside vendors, are key to timely completion of the evaluation without interruption.

(8) **Shipment instructions**. Complete instructions regarding the shipment of supplies, equipment, data, and/or patient specimens, if applicable, from the evaluation site to the sponsor. It is recommended, if commercial couriers are used for this purpose, that preprinted courier forms be provided to the evaluation site. If specimens are to be shipped, instructions must include the proper labeling and packaging of these specimens in compliance with good laboratory practices and governmental regulations.

(9) **Safety**. A complete analysis of any possible safety hazards which could confront the clinical investigator, including those associated with specimen collection/handling, reagent preparation/testing, or instrument/equipment operation. Refer to NCCLS document M29–*Protection of Laboratory Workers from Occupationally Acquired Infections* for additional information.

(10) **Environment**. Instructions for the safe and environmentally sound handling of reagent waste, residual patient specimens, packaging, reagent vapors/aerosols, or other disposable supplies associated with the clinical evaluation.

(11) **Responsibilities**. A description of responsibilities of each participant: the monitor, evaluation investigator, and additional personnel involved with the performance of the clinical evaluation.

Other items to include in the investigator's manual include:

- key site personnel *curriculum vitae;*

- monitor site visit log;
- subject enrollment list form;
- correspondence log (includes phone calls, letters, faxes, and email);
- laboratory certification; and
- investigational materials inventory accountability log.

3.2 Ethical Considerations

The protection of human subjects is critical to the conduct of medical research worldwide. Each evaluation investigator must obtain approval from the institutional review board (IRB) or the independent ethics committee (IEC), groups that are responsible for the implementation and oversight of methods to ensure that protections are appropriate and adequate.

To adequately protect the subjects involved in medical research, the investigator should consider: risk analysis and management; institutional review board approval; informed consent; linkage and patient confidentiality; and use of data in patient management. These issues are discussed in further detail below.

3.2.1 Risk Analysis and Management

Risks to the subjects must be minimized whether or not the evaluation requires review by an institutional review board. Risks may include physical harm, psychological harm, and violation of privacy. Any possible risks must be outweighed by the benefits from participation in the evaluation. These benefits might include improved medical treatment resulting from the evaluation, or possible benefits to society or future patients resulting from the introduction of a new assay or a new application of an existing assay.

3.2.2 Institutional Review Board

An institutional review board (IRB) is a group formally designated by an institution to review, approve, and monitor any research involving human subjects. Outside of the U.S., an IRB is often called an "ethics committee." Although most institutions have their own IRB, in some cases an investigator or sponsor may find it appropriate or necessary to contact a commercial IRB for evaluation review. Commercial IRBs generally charge a fee for their services. The purpose of the review board is to protect the welfare and rights of human subjects in any evaluation.

In reviewing any evaluation, the IRB will consider the scientific merit (including sound scientific design), the ratio of the risks to the benefits, and any bias in subject selection. The IRB will also consider the balance between ethical and social benefits and is responsible for the protection of any vulnerable populations (e.g., children, mentally infirm, prisoners).

Sufficient time should be allowed in the project plan for IRB review. Some protocol reviews can take up to three months depending on the issues raised by the IRB. The clinical evaluation protocol and the patient informed consent form are prepared by the sponsor and principal investigator. The IRB may request that a specific format be used for submitting evaluation information. The principal investigator submits these evaluation documents to the IRB for review. The IRB may waive the requirement for review, or conduct a full board review. The IRB determines the requirement for patient consent and will approve the specific language in the patient consent form. An original copy of the patient consent form, if required for the evaluation, must bear a dated stamp from the IRB. The portion of clinical evaluation using subjects or subject specimens cannot begin prior to receipt of an IRB waiver or approval documentation.

3.2.3 Informed Consent

If indicated as a requirement by the IRB or the IEC, the investigator must ensure that a patient consent form is signed by each evaluation subject. The consent form needs to be written using nontechnical language that is understandable to the subjects. In some cases the clinical investigators or their designee must provide oral explanations of all pertinent aspects of the evaluation. In signing the consent form, the subject acknowledges that his/her consent was given freely without coercion. The basic elements of informed consent include:[c]

(1) explanation of the purpose of the research, the duration of the evaluation, and a description of the procedures to be used (procedures are either experimental or a part of established medical practice);

(2) description of any reasonably foreseeable risks or discomforts to the patients;

(3) description of any benefits to the patients or to others;

(4) disclosure of appropriate alternative procedures;

(5) statement regarding the confidentiality of any records;

(6) explanation of any compensation and available treatments in the event of complications;

(7) contact name for answers to questions related to the research and patients' rights in the event of a research-related injury; and

(8) statement that participation is voluntary and may be discontinued at any time without penalty.

Many IRBs or IECs will not require informed consent for clinical evaluations using surplus specimens, provided that there is no link to the patient's identity. Informed consent can be required when clinical evaluations include, for example, access to patient medical records, specifications for patient treatment, or collection of specimens for the specific purpose of the evaluation.

3.2.4 Patient Confidentiality

Although clinical evaluations can require that the evaluation subject's specimen be linked to the patient's medical records, extreme care must be taken to protect the confidentiality of the patient. Identification numbers for subjects and/or clinical specimens can be recorded using a coded system to protect individual identity. Social security numbers and patient names should not be recorded as identifiers on data collection forms.

In clinical evaluations where a link between the specimen and patient is not required by the clinical protocol, information should be recorded such that subjects cannot be identified, directly or through identifiers linked to the subjects.

3.2.5 Use of Data in Patient Management

Use of the immunoassay data for patient management (e.g., treatment decisions) generated during a clinical evaluation of an *in vitro* diagnostic assay (IVD) is not generally accepted practice. This is particularly applicable to the use of IVD tests for which the performance characteristics have not yet been established. Exceptions include conditions in which the evaluation design requires disclosure of test results, or in which the researcher would be medically negligent by not informing the subjects. An

[c] In the U.S., refer to 21 CFR Part 50.25.

example of the need for disclosure of test results is found in the evaluation of volunteer blood donors with a new HIV-1/2 immunoassay. In this case, disclosure of positive results by the new assay is required to allow for donor deferral until donor suitability can be resolved.

3.3 Clinical Evaluation Protocol

The clinical protocol is the most important document associated with the clinical evaluation. It is the set of instructions that each investigator will follow to conduct the evaluation. It must be accurate, clinically correct, and written in clear, concise language. During protocol development, an evaluation sponsor can gain valuable insight into its effectiveness by reviewing the protocol with several key investigators. In addition, review by a biostatistician may be beneficial as the protocol is being developed. Prior to initiating the evaluation, it is important to ensure that the evaluation is feasible and meets the stated objectives. In addition to review by the investigators, it may be possible to review the protocol with the appropriate regulatory body. For example, in the U.S., the FDA will review draft clinical protocols for new assay systems requiring approval or clearance. The major elements of a clinical protocol are outlined below:

(1) **Table of Contents**. A well-indexed table of contents makes it easier for the laboratory staff to quickly find important information any time during the evaluation. All the information needed to perform the clinical evaluation can be found in this document.

(2) **Introduction.** Provides background information about the new assay, including proposed indications for use; describes clinical significance for which the assay is being evaluated; and describes or references previous research.

(3) **Description of the New Assay Method or Device**. Provides details on the method and procedure.

(4) **Objectives/End Points**. Specific, measurable objectives for the clinical evaluation that define the purpose of the evaluation, with reference to particular performance characteristics being evaluated, (e.g., clinical sensitivity, clinical specificity, etc.). Considers the need for well-defined clinical end points.

(5) **Design End Points**. Summary of the evaluation plan; defines measures for clinical diagnosis evaluation (if appropriate to evaluation design), duration of evaluation, type and estimated number of specimens or patients to be tested, statistical justification for sample size estimate, information on interpretation of results, plan for discrepancy resolution, and testing format, including test methods and specimen coding system.

(6) **Study Population**. Inclusion and exclusion criteria for patients and/or specimens; pertinent population demographics; logistics of recruiting and selecting subjects.

(7) **Description of Assay Methods**. Describes all other methods to be used in the clinical evaluation, including all methods for discrepancy analysis (e.g., new assay method, other immunoassays, confirmatory tests).

(8) **Site Training**. Describes training needed for the clinical evaluation, including any requirement for demonstrating proficiency with the new assay method, as well as any other methods to be used in the evaluation.

(9) **Specimens**. Acquisition, storage, handling, and shipping.

(10) **Warnings and Precautions**. Cautions for handling infectious material.

(11) **Limitations of the New Assay System**. Describes any stress conditions that might lead to product failure; lists any limitations of the new assay system.

(12) **Data Management**. Provides a description of methods of data collection, data collection forms, subject enrollment log, and procedures for transferring data to the sponsor.

(13) **Data Analysis**. Provides a description of statistical methods to be used for determining clinical and analytical performance characteristics.

(14) **Monitoring Schedule**. Identifies the evaluation monitor and primary method of communication with the site; indicates a schedule for monitoring visits.

(15) **Principal Investigator Responsibilities**. Some general obligations include: ensuring the investigation is conducted according to the protocol and the agreement with the study sponsor; protecting the rights of subjects; obtaining appropriate informed consent; controlling distribution of investigational devices; retaining records as specified by the protocol; and assuring that an IRB is informed for initial and continuing review of the study. Additional study specific responsibilities may be listed in the study protocol.

(16) **Materials**. Lists materials the sponsor will provide and those the investigator is expected to provide.

(17) **Key Contacts**. Lists contact information for both the evaluation site and the sponsor, complete with name, address, phone, FAX, and E-mail.

(18) **Appendixes**. May include: sample data collection forms, sample informed consent, draft product instructions, evaluation outline for quick reference, and flow chart for handling specimens through all possible result algorithms**.**

Additional elements to be considered for inclusion in the clinical protocol are:

- number of sites;
- masking (if appropriate);
- Subject discontinuation procedures;
- source documents and record retention;
- risk/benefit analysis, risk management;
- IRB and informed consent requirements; and
- subject confidentiality and subject ID log description.

(19) **Termination of evaluation**. May include closeout procedures and early rules/clauses for consideration of termination.

3.4 Clinical Evaluation Objectives

A successful clinical evaluation must have clearly established objectives and clinical end points. The objectives should be specific and measurable. For example, “to establish the clinical sensitivity and specificity of a new infectious disease assay” is too general. The objective is better stated in specific terms, identifying the clinical management question and the target population. Hence, a more complete statement of the objective might be “to estimate the clinical sensitivity and specificity of a new serum immunoassay for analyte X to be used to identify clinically important coronary artery disease (CAD) in women presenting with signs and/or symptoms suggestive of CAD.”

The objective is to distinguish those women with important CAD from those without; the target population would be symptomatic women, and the end point might be coronary angiography. A clinical evaluation may be designed with multiple objectives. A second objective might be to screen asymptomatic males to identify those with occult CAD. This is a separate question, requiring different subjects, and might even use a different specimen. Different clinical applications for a given test require separate evaluations, whether or not they are conducted simultaneously.

3.4.1 Clinical Feasibility/Pilot Evaluation

There may be a number of unknowns in the evaluation design that could be established with a pilot evaluation. Generally, a pilot evaluation has a similar design but may not have a statistically significant sample size relative to a full clinical evaluation. The outcome of the pilot evaluation can help to confirm the evaluation design for the larger evaluation. Some of the reasons for conducting a pilot evaluation include:

(1) to evaluate a new sample preparation method;

(2) to ensure that the evaluation design is feasible in the laboratory;

(3) to evaluate new data retrieval and data management methods;

(4) to better establish appropriate clinical end points;

(5) to confirm patient selection criteria, thus better defining the evaluation population;

(6) to evaluate the performance of the new assay with fresh (unfrozen) clinical specimens;

(7) to evaluate a new assay system in the laboratory setting (for instance, combining the reagents and instruments for the first time);

(8) to determine expected variability and means in assay values for different patient populations and/or proposed diagnostic cutoffs to better estimate the sample size for the clinical evaluation;

(9) to modify and optimize (improve) the evaluation protocol; and

(10) to validate the performance of the investigator's method or the use of the reference assay system.

3.5 Selection of Investigator and Evaluation Site

Selection of qualified investigators is an essential part of any successful clinical evaluation. The following is a list of attributes to look for when selecting an investigator and clinical evaluation site:

(1) The investigator has demonstrated expertise in the clinical field of interest.

(2) The investigator has sufficient time from other commitments for the evaluation.

(3) The investigator/institution is unconstrained by other commitments.

(4) Consider the laboratory's experience with particular technology associated with the new assay.

(5) Consider the laboratory's experience with other methods to be used during the clinical evaluation.

(6) The investigator has previous experience with clinical studies.

(7) The investigator/institution demonstrates willingness to participate.

(8) The investigator/staff is committed to complying with good clinical practice (GCP) standards.

(9) The investigator/institution understands the need for informed consent.

(10) IRB reviews are conducted on a timely basis.

(11) The investigator/staff are committed to following the evaluation protocol.

(12) The facilities are adequate), and there is secure storage for records and investigational product.

(13) Convenient access to patient records is provided.

(14) There is a well-organized approach to documentation and record keeping.

(15) Sufficiently qualified staff can be assigned to the evaluation, especially an evaluation coordinator.

(16) The staff turnover rate is low.

(17) The staff are given sufficient time from other duties to accomplish the evaluation.

(18) Patients/subjects who are willing to participate in research are accessible.

(19) The patient population/demographics are consistent with the evaluation design.

(20) Adequate number of subjects/specimens meeting inclusion/exclusions criteria exists, as well as the ability to enroll enough subjects consistent with evaluation timelines and total specimen targets.

A visit to the evaluation site prior to final selection of investigators can be valuable in evaluating the above criteria. A face-to-face interview with the investigator can sometimes be more revealing than a phone conversation. Good clinical practice standards refer to the following types of site visits: 1) a pre-evaluation visit to confirm staff and facilities are adequate; 2) an evaluation initiation visit, in order to train the site personnel regarding the protocol and evaluation requirements, as well as confirming that the new assay method is working well; 3) a visit to monitor evaluation progress and review raw data and patient charts; and 4) a close-out visit to conclude all evaluation activities and confirm that all evaluation documents are in order.

3.6 Evaluating Performance Characteristics

3.6.1 Clinical Performance

What is “clinical performance?” Clinical performance is the performance of a diagnostic test in a clinical context. Clearly defined objectives and end points naturally describe the intended diagnostic use of the assay in its medical context. This use indicates what the relevant performance characteristics under investigation are, and therefore, what should be evaluated. What is the clinical management issue? Who are the subjects relevant to this objective? What are the end points? An example might be as follows:

Objective: To identify, among individuals who are over a certain age and have positive results for fecal occult blood (FOB), those who are candidates for colonoscopic exam. Subjects: In this case, persons over the age of interest and having a positive FOB. End point: Presence or absence of cancer based on histopathologic examination of lesions identified and removed in a direct exam of the colon.

Assume an immunoassay for a serum marker for colon cancer is being evaluated for this purpose. The subjects have some probability of colon cancer as a result of their age and positive FOB. The issue is whether the test can provide new information to revise that pretest probability (prior probability) for each individual to a posttest probability of cancer. This posttest probability would be used to distinguish between those who are candidates for some further exam (such as colonoscopy) and those that need not have it. This ability to separate the subjects is the clinical performance of interest.

Possible uses of an assay include: (1) screening for disease or conditions in the absence of indications; (2) pursuing a differential diagnostic workup in response to signs/symptoms or some other clinical indications; (3) following a known condition to determine a change in clinical status (with or without interventions); (4) choosing among intervention options; or (5) assessing the status of an individual, such as whether they have been immunized or exposed to some infectious agent. In each of the possible applications, a specific question is being addressed about a defined target group of subjects, in an attempt to make some clinically useful distinction among the subjects. The ability of the assay to make this distinction, that is, to discriminate among the relevant subgroups (for example, healthy versus diabetic; normal risk versus increased risk; prostate cancer versus benign hypertrophy in subjects with difficulty urinating; relapse of cancer following therapy; subjects with estrogen receptor positive versus negative breast cancers; potential blood donors with and without hepatitis) is often termed "clinical accuracy" and is the performance characteristic at issue.

How can this discrimination ability be described and evaluated? One common approach is by determining clinical sensitivity and specificity. Generally sensitivity and specificity vary as the decision threshold (cutoff) changes, and they are reciprocal. As one increases, the other decreases. A test, then, exhibits a spectrum of trade-offs between sensitivity and specificity as the cutoff changes. For continuous variables, the most comprehensive description of this spectrum of clinical sensitivity and specificity is provided by the receiver operating characteristic (ROC) analysis, because it examines all possible decision thresholds and all the corresponding combinations of sensitivity and specificity. This makes it a global assessment. It is independent of prevalence (prior probability) as well. The ROC plot itself is a qualitative representation of clinical sensitivity and specificity; the area under the ROC plot is a quantitative one. (See the most current edition of NCCLS document GP10—*Assessment of the Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plots*.) In the example mentioned above, ROC of the results for a new serum marker of colon cancer would indicate the ability of this new test to discriminate between those individuals who were found to have an unknown cancer and those who did not. The ROC data could also be used to compare this new marker to an existing one. And further, the ROC data could be combined with other relevant data on prevalence and on the cost of false results to evaluate and select particular decision thresholds for clinical use. For assays with semiquantitative results such as "negative," 1+, 2+, etc., ROC analysis can be employed, though the smaller number of different results will yield plots with fewer points.

Once particular decision thresholds are identified, performance characteristics, such as positive and negative predictive value (PV), may be useful descriptors of the meaning of individual test results. PV analysis, unlike ROC analysis, is prevalence-dependent. ROC describes how well the test discriminates between two categories of subjects, while PV describes the probable meaning of normal or abnormal results once the cutoff and prevalence are defined.

For purely qualitative data, test results expressed simply as "positive" or "negative" in 2 x 2 tables as illustrated below, may be used to characterize the clinical performance of each test.

		# Subjects	
		Affected	Unaffected
# Test Results	Positive	A	B
	Negative	C	D

Using these tables, from two or more different tests, assays can be compared to one another using the appropriate statistics.

3.6.2 Analytical Performance

The analytical performance characteristics of a diagnostic test should be well defined before instituting a clinical evaluation. The clinical evaluation phase should include verification components for maintenance of the analytical performance characteristics. This verification of performance characteristics is also an important series of measurements to ensure that the analytical performance characteristics of a particular test are reproducible at different test sites and with different staff performing and interpreting the test. For many IVDs, the analytical performance characteristics may be established by the manufacturer or a central laboratory/researcher within an institution. In these instances, the analytical performance characteristics, including traceability, trueness, and precision (repeatability and reproducibility), must be established/verified at each clinical evaluation site to ensure the comparability of clinical test data prior to testing clinical specimens. The manufacturer or central laboratory should provide each test site with selected control materials that can verify the trueness and reproducibility of analytical performance.

3.6.3 Statistical Power and Sample Size Considerations

Determining the sample size for a clinical evaluation may require the assistance of a biostatistician to define the evaluation objectives, selection of evaluation populations, and statistical methods. If the intent of the clinical evaluation is to determine test efficacy or compare different tests using ROC plots, then the sample size may require sufficient numbers to provide a valid estimate of the ROC plot for the IVD. The clinical performance of an IVD may be characterized by determination of clinical sensitivity and specificity, as well as the confidence intervals associated with each of these performance characteristics at selected decision thresholds. Since the confidence intervals for either the sensitivity or the specificity will decrease as the sample size increases, the sample size calculations for a clinical evaluation may also include prior considerations about the desired confidence intervals for sensitivity and specificity.

Sample size considerations will also depend on the intended use of the test, the evaluation population(s), and the prevalence of disease in the evaluation population. For many diseases, a clinical evaluation may require multiple sites to provide sufficient numbers of diseased individuals and to analyze test data in different populations.

In studies with the intent to determine reference values and reference intervals for quantitative tests, the sample size is determined by the reference interval the investigator wishes to present. More samples are needed if a wider reference interval estimate (such as central 95% as compared to 90%) is targeted. In any case, a minimum of 120 subjects is recommended for such a study. For more information on determining the sample size in a reference intervals study, see NCCLS document C28—*How to Define and Determine Reference Intervals in the Clinical Laboratory.*

3.6.4 Methods of Statistical Analysis (Confidence Intervals and Considerations for Pooling Data)

Techniques for plotting ROC curves, calculating areas under the plots, and determining relevant confidence intervals have been described.[3,4] Computer software is available. In order to pool data from multiple evaluation sites, the evaluation must be designed and executed to ensure consistency from site to site in all aspects, particularly selection of subjects, sample collection and storage, and analytical techniques. This consistency must be verified before the data from the various sites are actually merged. Consultation with biostatisticians is recommended while the evaluation is being designed.

Any subjects, samples, or data which are lost, deleted, or otherwise do not conform to the protocol should be accounted for and described.

3.6.5 Reproducibility Testing

Reproducibility testing may be assessed on the basis of lot-to-lot, day-to-day, run-to-run, and technologist-to-technologist variation of the new assay method. This portion of the evaluation is generally conducted with a panel of specimens provided to each site participating in the evaluation (See the most current version of NCCLS document EP5—*Evaluation of Precision Performance of Clinical Chemistry Devices.*

3.7 Classification of Subjects

For most clinical evaluations that require patient information, definitive diagnoses of enrolled subjects are required for accurate classification. Ensuring that each subject meets certain criteria may require extensive workup and/or access to medical records. Clear classification of some subjects may be difficult or impossible (see Section 3.7.2). Excluding subjects with equivocal diagnoses from the data analysis must be explicitly acknowledged. Such exclusions introduce bias in the estimates of clinical sensitivity and specificity.

3.7.1 Algorithm

It is important that the actual state of health of each subject be established as definitively as possible and without knowledge of the test results under evaluation. It is this definitive classification against which the classification provided by the assays under evaluation will be assessed. In deriving this algorithm for classification, it is desirable to go beyond routine means of clinical classification and instead design as thorough a clinical analysis as feasible. This analysis may include, for example, histopathology, state-of-the-art imaging, nonroutine laboratory testing, surgical information, autopsy data, and short- or long-term clinical outcome. The assays under evaluation should not be included in the algorithm for establishing the definitive classification of subjects.

3.7.2 Strengths and Limitations of Classification

The strongest evaluation will be based on definitive classification of the subjects. Any error in the primary classification will result in distortion (either upward or downward) of the apparent clinical accuracy of the assays under evaluation.[3,5,6] This is why a rigorous workup is desirable. It is not unusual, however, that a truly definitive classification cannot be established for a particular disease state (myocardial infarction, for example), or for particular individuals with equivocal findings. Alternative approaches have been reported and discussed.[3,7,8,9] In the case of the former problem, the validity of the entire evaluation is limited and should be recognized as such. Sometimes, clinical follow-up and determination of outcome can provide a useful alternative to a real-time diagnosis or categorization. In fact, outcome may be the preferred means to classify subjects. For example, in the case of a prenatal assay for a genetic abnormality such as Down's syndrome, waiting until parturition and then examining the newborn child, rather than trying to classify the fetus *in utero* would provide a truly definitive

classification. Likewise, waiting to observe the behavior of a tumor may provide a more relevant clinical classification which may be more definitive than imaging or even histopathology.

3.7.3 Analysis of Discrepant Specimens

When a predicate device (a currently available diagnostic test) is used in place of a definitive diagnosis to assess the clinical performance of a new test (or new clinical application), discrepancies may occur between or among the results of the various test devices. This is anticipated, since rarely would results match perfectly among two or more devices. Even when results do match perfectly, both the predicate and the new test may be incorrect. It is often assumed that the percentage of incorrect but matching results is zero or small. Even if this assumption is not true, comparative information about the clinical accuracy of the predicate device versus the new test can still be obtained by investigating only the discrepant specimens using a definitive diagnosis to see whether the new test or the predicate is correct. This investigation provides information on whether the predicate or the new test makes fewer classification errors, and the results from the discrepancy analysis can be provided to the user. However, the results should not be provided in the performance calculations, since the investigation does not provide data needed to calculate statistically unbiased estimates of performance for either test. If the additional testing used to further characterize discrepant specimens is not 100% accurate, then discrepancy resolution may not be useful at all. The only information obtained is whether the predicate and the new test agree, but it provides no information about the clinical accuracy of the two tests. It is appropriate to calculate performance characteristics if the protocol involves a definitive diagnosis on all specimens, or on all discrepant specimens and a random sample of specimens where initial results agreed. The latter requires statistical and practical planning in the protocol stage, and may not be simple or cost effective. It is recommended that planning for discrepant analysis be included in the study design. A scheme for evaluating the relevant subjects to determine their actual clinical classification can be established in the study design and then implemented after the results are obtained.

The frequency of discrepancies can be described by using a 2 x 2 table as shown below.

		# Results Test 2		
		Positive	Negative	
# Results Test 1	Positive	A	B	B and C are discrepant
	Negative	C	D	

Cases falling in the lower left and upper right boxes are discrepant. Note that changing the decision threshold, or cutoff, for any given assay will change the numbers appearing in the boxes. Thus discrepancy results can increase or decrease as the decision thresholds vary. While varying the cutoff and examining the change in the discrepancies can provide insight into the nature of the discrepancies, the actual choice of the new test cutoff should be made prior to the start of the comparison study.

3.8 Evaluation Population

3.8.1 Establish Clinically Relevant Population

Once the clinical question being addressed by the clinical evaluation is clearly and fully defined (see Section 4.1), the targeted population will be evident. Nevertheless, an explicit and detailed description of this relevant clinical population may be developed, so that relevant samples can be defined by rigorous,

unambiguous inclusion/exclusion criteria. Careful attention is required to ensure an appropriate spectrum of individuals.[3,5,6,10,11] The subjects actually selected for evaluation in a clinical evaluation should be representative of that larger subject population for whom the test is ultimately intended. Selection procedures which use only the most convenient or readily accessible subjects are likely to introduce bias into the evaluation by obtaining a non-representative sample. For example, a selection bias may be created by enrolling employees, only those patients seen early in the day on weekdays, only those subjects for whom the laboratory has leftover samples, or only those subjects whose true classification is clearly and easily established. The conclusions of such an evaluation may not be applicable to the larger population.

As mentioned above and in the literature, healthy controls or other so-called "normals" may not be relevant to the evaluation, unless the clinical question being tested by the evaluation involves using the test for screening such apparently healthy individuals. If this were the case, then some defined group of healthy or asymptomatic individuals would comprise the entire evaluation sample.

3.8.2 Subject Exclusion/Inclusion Criteria

The inclusion/exclusion criteria uniquely characterize the evaluation population and relate this population to the intended use of the test. Care should be taken to avoid biases in the criteria and in the actual selection process.[10,11] Criteria for selection of subjects should be clearly established before selection begins and then adhered to strictly in conducting the evaluation. Examples of inclusion/exclusion criteria include medication use, presence of a disease state or stage, and fasting or dietary requirements. In addition to patient criteria, immunoassay evaluations must also consider specimen criteria, such as storage conditions, time from procurement to testing, and specimen volume and quality. Carefully designed and written criteria are more likely to provide data that can be pooled from multiple sites for statistical analysis. All sites involved must be able and willing to agree to the criteria and to operate in strict conformity.

3.8.3 Patient History Documentation

After establishing subject information requirements for the evaluation, procedures for collecting and recording the data should be established and documented. The sponsor should consider that access to patient charts and data may be facilitated if the physician evaluating the patient is the principle investigator or coinvestigator. Depending upon institutional or sponsor requirements, each documented history must be signed by the clinical personnel taking the information. All sites should agree to conform to these procedures and maintain confidentiality of all data.

3.9 Considerations for Masking

The use of masking, formerly termed "blinding," may be necessary to prevent bias that may be introduced through knowledge of results of an ongoing evaluation. Evaluator bias is a type of investigator bias in which the person interpreting or recording test results may inadvertently bias the measurements through subtle changes that may favor the test method being evaluated. When evaluating a test method, the investigators should not have access to either the clinical information or the reference test method result. The test method results should be recorded for analyses and comparison with the reference methods and clinical data after completion of the evaluation. Each patient specimen should be separately coded so that only the principle investigator or sponsor can link patient clinical information; however, the clinical investigator or the sponsor must be able to link results from the test undergoing evaluation with reference test results. Although the use of different test personnel and instrumentation with objective quantitation would seem to minimize bias, there are many sources of bias that cannot be retrospectively measured or controlled. The use of coded patient samples, with no access to patient identifiers or patient clinical data, is the only method to ensure that evaluator bias has not altered the clinical evaluation data.

4 Conducting the Clinical Evaluation

4.1 Monitoring Clinical Evaluations

The responsibilities of the monitor are to ensure the safety, rights, and welfare of the patients in the evaluation, to ensure the quality and integrity of the evaluation data, and to ensure adherence to the evaluation protocol. It is recommended that the clinical monitor be an individual who possesses the education, experience, and communication skills appropriate to monitor the progress of the clinical evaluation. In the case of a clinical evaluation sponsored by a manufacturer, the monitor is typically identified within the sponsoring organization. In the case of an evaluation initiated by an end-user laboratory, the monitor may be identified within the end user's organization. Alternatively, a contract research organization (CRO) may be contracted to monitor an evaluation. (In the U.S., refer to the FDA *Guideline for the Monitoring of Clinical Investigations*; 1988). The duties of the clinical monitor include, but are not limited to:

(1) assay procedure training for clinical laboratory staff who will perform the testing;

(2) providing training on the study protocol and study forms for all appropriate personnel, including the study coordinator, principle investigator, study nurses, and clinical laboratory staff;

(3) acting as the primary communication link between the sponsor and the evaluation site;

(4) verifying records/data;

(5) verifying forms completion;

(6) verifying compliance with the evaluation protocol;

(7) reviewing IRB correspondence;

(8) reviewing original patient informed consent forms;

(9) maintaining device accountability;

(10) documenting monitoring activities, such as phone logs and site visit reports;

(11) conducting site visits; and

(12) ensuring that sites have sufficient supplies of forms, evaluation materials, and investigational kits and reagents to complete the evaluation.

Training the clinical laboratory staff who perform the testing is an important aspect of the clinical evaluation. Verification that training was effective can be easily accomplished by providing a panel of test specimens having known values. Using these panels, laboratory personnel from each site can demonstrate proficiency with the new assay method prior to commencing the evaluation. This can be an effective test of the training materials and new assay instructions, as well as the capability of the laboratory personnel. If identical performance of a complex method is important to the evaluation design, demonstration of proficiency at all evaluation sites is essential prior to commencing the evaluation. It is well worth incurring a short delay while additional training is provided or personnel are reassigned to ensure the quality of the data.

Close monitoring of the data as it is collected can help to quickly identify problems which can then be addressed in a timely manner. Periodic site-monitoring visits may be required to manage the evaluation.

For simple evaluation designs, phone communication may be sufficient. For studies of short duration, daily phone calls may be necessary; for longer studies, weekly calls may be more appropriate. Electronic transfer of data can facilitate monitoring from a distance. Alternatively, data sent to the sponsor by fax on a regular schedule may also suffice.

It is important to keep a record of all phone and on-site monitor communications. The record should include information reviewed, findings, conclusions, and action taken to correct any deficiencies. A pre-evaluation visit is conducted to ensure that the investigator and the laboratory facility qualify for the evaluation. Monitoring visits during the evaluation may be necessary to ensure data integrity and to verify protocol compliance. A closeout visit at the end of the evaluation is essential to complete the account of all investigational devices, and to ensure that all evaluation records are in order.

4.2 Management of Database

The evaluation sponsor is responsible for designing, verifying, and maintaining a database management system. Critical to the successful completion of an evaluation, the database management system design must meet the requirements of the clinical protocol for ensuring accurate and reliable data. Optimally, the design should be verified well in advance of the start of testing to allow for modifications. A clinical feasibility or pilot evaluation may serve as a means of verifying the management system. The system should be well documented with information that may include:

(1) the names and roles of each site's principal investigators, coinvestigators, and technical staff;

(2) a schedule of reagent/kit shipments and lot numbers;

(3) the mode of data transmission, such as paper copies, fax, floppy disk, or by modem;

(4) an example and explanation of a data reporting form and case report form;

(5) a description of data entry into the sponsor's database and means of ensuring clerical accuracy;

(6) a brief description of data availability for statistical analysis;

(7) a means of identifying and analyzing discrepant results, as well as missing data;

(8) procedures for conducting data audits and for general compliance;

(9) general troubleshooting information;

(10) procedures for data review; and

(11) a description of procedures for archiving data.

In addition to defining these activities, the sponsor should clarify the general flow of data. For example, most clinical sites will transfer their data to the monitor, who ensures delivery to the database manager. Once the data is entered into the database, the manager may submit it for statistical analysis. The statistician may then send his/her analysis to the monitor for incorporation into a regulatory submission, internal documentation, or publication. Consideration must also be given to the site procedures for documenting supporting data, such as the clinical history or ongoing testing, including x-rays, body scans, or biopsies.

4.3 Quality Assurance of Data Integrity

Creating conditions to minimize mistakes should be a primary concern of the sponsor and the investigator. Responsibility for efforts to reduce mistakes lies with the sponsor and with the investigative team. The clinical monitor should plan an ongoing data review for 100% accuracy of records and total compliance by the site personnel. This review is considered part of the monitoring process. Data audits are conducted by sampling the evaluation population. Typically, audits involve 5 to 10% randomly selected records. If problem records are detected, an additional 10% may be audited to determine the extent of the errors. Ideally, audits are conducted by internal QA data auditors or by external contractors. Because these resources may not be available, audits may be completed by the monitor or his/her designee.

An initial step in reducing mistakes is ensuring adequate training of the investigative team. Prior to the start of the evaluation, the technical personnel at the site should be trained in all aspects of testing using the new immunoassay and any comparative tests that may be part of the evaluation design. The site should be thoroughly familiar with the evaluation protocol and inclusion/exclusion criteria for subjects or specimens. In addition, the site personnel should be instructed on the proper means of using the data forms or other data applications, and should have a thorough review of troubleshooting. The monitor should develop a means of verifying that the site personnel have received this training and have demonstrated a proficiency in testing and transferring data.

An audit of the database management system provides the sponsor with a direct estimate of the database quality. Assuming a 100% data entry check is completed as data is transferred to the sponsor, audits or data queries can be scheduled as often as needed to complete the evaluation. The audits should follow the data flow, with special attention given to the data transfer points. For example, the auditor should check that any information extracted from the patient's chart (source document) is accurate and that data reporting forms accurately reflect instrument printer results. In addition, the auditor should review the handling of warning flags, missing data, and outliers, as well as any changes in data entries. All data queries that are detected during an audit should be resolved and documented.

4.4 Retention of Records

Data from clinical evaluations shall be retained in accordance with regulations in countries where the product is being registered or used.[d] For those laboratories conducting a methods comparison for a new assay or a new clinical application, the retention time begins when the assay is placed into clinical practice. Records should be retained in an easily accessible location that would allow for review by agency or laboratory surveyor personnel.

For any evaluation, the records may take the form of hard copy (paper) documents and report forms, computer disks, computer mainframe tapes, or other forms of electronic transmission. The monitor is responsible for a plan that ensures retention of original data that cannot be modified without clear notation and justification.

5 Analysis of Clinical Evaluation Data

The analysis should clearly state the hypothesis (clinical issue) tested, the statistical tests used to describe the data obtained, and the assumptions behind the tests. All statistical procedures should be referenced.

[d] In the case of assays requiring agency review, such as in the U.S., the investigator and sponsor are required to retain records for two years from the time of regulatory clearance or approval or from the time withdrawn by the sponsor.

5.1 Performance of Statistical Tests

Processing of data should be determined in advance, including the statistics to be calculated and the method of calculation (see Sections 3.6.1 and 3.7.3). Computer software is generally available for analyzing data generated in the evaluation. Appropriate statistical approaches should be identified at the design stage through consultation with a statistician, and relevant software should then be selected, implemented, and validated. Before pooling data from multiple sites or multiple evaluations, it is important to establish the criteria for pooling data.

5.1.1 Continuous Quantitative Data

Tests providing continuous data should be evaluated (and compared to one another) by ROC analysis as already mentioned (see Section 3.6.1). Complete plots of false-positive fraction versus true-positive fraction (or an equivalent alternative plot, such as true-negative fraction versus true-positive fraction) should be generated for each test, along with confidence intervals. Areas under the plot, along with their confidence intervals, should also be obtained. The significance of any differences observed among plots and/or among areas under the plots can be determined.

5.1.2 Discontinuous (i.e., discrete or ratings) Quantitative Data

For discontinuous data with more than a few classes of results, ROC analysis can be used to evaluate results and to compare tests, though some variations may be required.

5.1.3 Qualitative Data

As discussed in Section 3.6.1, qualitative data is best characterized by application to a 2 x 2 discrepancy table in which the positive or negative test results are compared to reference results. The reference may be represented by the actual "affected status" of the subject or by the positive or negative results of a predicate assay. Because a perfect reference capable of identifying true-positive specimens and true-negative specimens is rarely available, the investigator should have a good understanding of the reference assay true performance. In some cases, a more precise comparison may be made using two references, where their combined efficiencies provide greater clinical accuracy.

ROC analysis may be appropriate if the positive or negative test result is dependent upon a measured signal that is compared to a preassigned cutoff.

5.2 Documentation of Performance Characteristics

Major outcomes of a well-planned and well-executed immunoassay clinical evaluation are the data used to calculate performance characteristics of the assay under evaluation. Performance characteristics are generated for inclusion in the package insert which accompanies the product as it is sold to the end-user laboratory or, if for in-house use only, to appear in assay documentation records. These characteristics represent the performance that is to be expected when testing a population(s) similar to that evaluated during the assay clinical evaluation.

Ideally, clinical sensitivity and clinical specificity should be established by comparing the new test to a definitive patient classification (diagnosis) or other measure of clinical or operational truth, i.e., use of a well-established diagnostic reference method or combination of methods. When estimates of clinical sensitivity or, specificity are not possible, agreement of the new device in comparison to a predicate or comparator assay should be clearly reported. For each performance characteristic that is claimed, there should be a calculation of the exact 95% confidence interval. In addition, a description of the population(s) evaluated and the number of specimens tested, as well as a complete description of the discrepant resolution algorithm, should be included in any data summary.

Intra-assay and interassay measures of repeatability and overall reproducibility should be presented when describing the immunoassay performance characteristics. Where appropriate, the presentation of performance data in tabular format is recommended.

5.3 Clinical Evaluation Summary

The format of the clinical evaluation summary will be dictated by the purpose for writing it. A clinical evaluation summary may be intended for regulatory submissions, literature publications, clinical laboratory reports, or inclusion into manufacturer product inserts.

A summary of studies for regulatory submission might be expected to include:

(1) Clinical studies:

- Evaluation objectives.
- Overview (a summary of the evaluation design and evaluation plan).
- Clinical evaluation sites and qualification.
- Interpretation of results.
- Testing algorithms (detailed plan for resolving discrepant results for a methods comparison evaluation).
- Methods of statistical analysis (including statistical justification for sample size).
- Evaluation summary. In this portion of the report, the data to address each evaluation objective is described separately, along with a description of the patient population (inclusion/exclusion criteria and demographics). Headings for this section might include: clinical sensitivity, clinical specificity, calibration, reproducibility testing, etc.
- Conclusions.

(2) Preclinical studies (nonclinical):

- Interfering substances/cross-reactivity.
- Factors influencing specimen collection, handling, and processing.
- Specimen requirements and storage conditions.
- Analytical performance characteristics.
- Microbial contamination.
- Antibody/antigen characteristics.

A document describing the assay performance characteristics may include:

(1) Intended uses of the test.
(2) Summary and explanation of the test.
(3) Principles of the procedure.
(4) Warnings and precautions. (May include cautions for handling infectious materials and/or special handling for specific chemical components, or a statement “for investigational use only” or “for research use only”).
(5) Materials provided.
(6) Materials required but not provided.
(7) Description of the procedure.
(8) Quality control.
(9) Limitations of the procedure. (Describe any stress conditions that might lead to product failure; list any limitations of the new assay system).
(10) Expected results and their interpretation (if applicable).
(11) Performance characteristics:

- Clinical performance (e.g., clinical sensitivity, clinical specificity, NPV, PPV).
- Analytical performance (minimal detectable concentration, analytical specificity, calibration, reproducibility, etc).

(12) References (bibliography).

References

[1] PrEN ISO 17511. (Draft) In vitro diagnostic medical devices—Measurement of quantities in biological samples—Metrological traceability of values assigned to calibrators and control materials. Geneva: International Organization for Standardization; 2000.

[2] Tietz NW, ed. *Clinical Guide to Laboratory Tests*. 3rd ed. Philadelphia: WB Saunders; 1995:358.

[3] Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental tool in clinical medicine. *Clin Chem.* 1993;39:561-577.

[4] McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver-operating characteristic (ROC) curves. *Med Decis Making.* 1984;4:137-150.

[5] Robertson EA, Zweig MH, Van Steirteghem AC. Evaluating the clinical efficacy of laboratory tests. *Am J Clin Path.* 1983;79:78-86.

[6] Zweig MH, Robertson EA. Why we need better test evaluations. *Clin Chem.* 1982;28:1272-1276.

[7] Valenstein PN. Evaluating diagnostic tests with imperfect standards. *Am J Clin Pathol.* 1990;93:252-258.

[8] Revesz G, Kundel HL, Bonitatibus M. The effect of verification on the assessment of imaging techniques. *Invest Radiol.* 1983;18:194-198.

[9] Henkelman RM, Kay I, Bronskill MJ. Receiver operator characteristic (ROC) analysis without truth. *Med Decis Making.* 1990;10:24-29.

[10] Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med.* 1978;299:926-930.

[11] Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. *JAMA.* 1995;274:645-651.

Additional References

Bross I. Misclassification in 2 x 2 tables. *Biometrics.* 1954;10:478-486.

Hagdu A. Bias in the evaluation of DNA-amplification tests for detecting *Chlamydia trachomatis*. *Stat Med.* 1997;16:1391-1399.

Lipman HB, Astles JR. Quantifying the bias associated with use of discrepant analysis. *Clin Chem.* 1998;44;1:108-115.

Miller WC. Bias in discrepant analysis: When two wrongs don't make a right. *J Clin Epidemiol.* 1998;51:219-231.

Peters T, Westgard JO. Evaluation of methods. In: Tietz. *Textbook of Clinical Chemistry*. Burtis and Ashwood, eds. 1995;225-237.

Galen RSA. Gambino SR. *Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses*. New York: Wiley; 1975.

International and European Standards

PrEN 13612. Performance evaluation of in vitro diagnostic medical devices. Brussels: European Committee for Standardization (CEN); 1999.

Food and Drug Administration (FDA) Guidelines

"Guideline for the Monitoring of Clinical Investigations," published January 1988.
PMA Review Statistical Checklist

"Points to Consider Guidance Document on Assayed and Unassayed Quality Control Material," draft February 1999.

Federal Regulations

FDA Center for Research and Radiological Health (CDRH) Internet web site: www.fda.gov/cdrh/
21 CFR 50 - Protection of Human Subjects
21 CFR 54 – Financial Disclosure by Clinical Investigators
21 CFR 56 - Institutional Review Boards
21 CFR 812 - Investigational Device Exemptions (includes responsibilities for sponsors, monitors, and investigators)
21 CFR 814 - Premarket Approval (PMA) of Medical Devices
International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. ICH Steering Committee. May 1, 1996.

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 Subcommittee Responses

I/LA21-P: *Clinical Evaluation of Immunoassays; Proposed Guideline*

General

1. The Introduction and Section 2 should be made consistent with the concepts and requirements of prEN/ISO 17511 (Metrological traceability of values assigned to calibrators and control materials).

   Comments were limited to those items involving inconsistencies with terms and definitions used in this NCCLS document. These terms and definitions are also currently inconsistent with well-assessed ISO Standards, like VIM (1993), GUM (1993), ISO 3534-1 (1993), ISO 5725-1 (1994), and others which are well known at ISO/TC 212 level. In addition, EDMA has prepared a terminology draft document on the subject based on the definitions reported in ISO and CEN standards and in some relevant publications presently available.

- **Sections 2.1 and 2.1.2 have been revised to be consistent with the traceability concepts in prEN ISO 17511 and, where appropriate, the subcommittee has modified terminology to be consistent with international use. Please see the "Note on Terminology" in the Foreword and Section 1.2. Also see the responses to Comments 21, 22, and 36.**

2. Should it be necessary to compare different vendors' antigens and/or antisera in the preclinical stages of establishment of analytical performance of an immunoassay (i.e., antisera may have been produced slightly differently, or to different epitopes of the analyte)?

- **This detail was not intended to be part of the scope or intent of this document.**

3. During the course of the clinical evaluation, are patients' symptoms followed or communicated to the investigator (i.e., remission of disease state; exuberation of symptoms; development of side effects from therapies, etc.)?

- **Where appropriate, patients' signs and symptoms are followed for the study design. For example, an integral part of the evaluation of the assay could be to predict clinical end points such as the response to therapy, or remission or exacerbation of disease.**

4. In the case of discrepant specimen results between the existing method and the new test, could the third test results be biased if there are limited methodologies to choose from for the analyte in question (i.e., it may have already been evaluated against the existing method)?

- **Section 3.7.3 has been revised for clarification.**

5. Reference to commercial controls does not consider unassayed controls or controls that don't have values for specific methods. Did you consider providing information about the utility of these?

- **As stated in Section 2.1.2, if using unassayed controls, developers should establish acceptability limits prior to beginning the evaluation. The subcommittee has provided an additional reference to support the text.**

6. Discussion on specimen stability and storage does not mention doing studies that evaluate the intended use/environment for the test, e.g., specimens for tests intended for use in an outpatient testing environment should have extended stability studies to sufficiently cover the typical amount of time required for specimen transport. This may be particularly important for laboratories that receive large numbers of specimens from off-site.

- **There is adequate discussion of this topic in Sections 2.1.2, 2.2.1, 2.2.2 and 3.1; but the onus of responsibility to address, establish, and explain any caveats for the assay (e.g., testing environment, stability requirements, transport, etc.) lies with the assay developer or the manufacturer.**

7. The description of reportable range does not address assays with single point calibration. Should it?

- **See the response to Comment 23.**

8. The daily work organization seems a little sketchy. Have you considered adding other important details, such as run size, calibration and control frequency, etc.?

- **This document provides a representative list of daily tasks. Each study is expected to have a unique list, which should be delineated in the study protocol.**

9. This document, although a good discussion document, does not address specific issues regarding how to carry out clinicals. Perhaps more detail regarding evaluation that is needed for quantitative vs. qualitative assays or screening vs. diagnostic assay would be appropriate.

- **This is not the intent of the document. The subcommittee recommends that the reader refer to other reference materials such as NCCLS document GP10—*Assessment of the Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plots*; 510K summaries; PMA summaries of safety and effectiveness; manufacturers' product package inserts; ICH guidelines with harmonized tripartite guidelines; or literature references for specific analytes and intended uses.**

10. There is minimal discussion regarding the depth of evaluation that must be conducted for different assays with various intended uses. Perhaps a table could be developed that would give specific analytes and the relative degree of evaluation that must occur, e.g., what must be done for a TSH assay vs. a cancer marker assay.

- **The subcommittee refers readers to NCCLS documents GP10—*Assessment of the Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plots* and I/LA18—*Specifications for Immunological Testing for Infectious Disease* (see Related NCCLS Publications).**

Foreword

11. Clinical evaluation, clinical study, clinical trial. This paper should distinguish carefully between "performance evaluation" and "clinical evaluation." In Europe the IVD directive requires performance evaluation of all commercial medical devices including immunoassay kits. Performance

evaluation is further described in prEN 13612. "Clinical study" or "clinical trial" is not required and is not necessary to ensure safe and proper use.

Whether an immunoassay has clinical relevance in medical utility is a question of the practice of laboratory medicine and not a question of product safety. Even if an immunoassay performs exactly, its clinical performance will vary from study to study because of the heterogeneity of patient samples and the fact that the assay can only be defined by the reactivity of the antibody used.

- **Please see the Note on Terminology in the Foreword and the Definitions section. Also, prEN 13162 has been added as an additional reference.**

Section 1.2

12. The first definition given for analytical sensitivity. "Quantitative testing" is the correct ISO definition for analytical sensitivity (i.e., slope of the calibration curve). By contrast, the second definition for "qualitative testing" cannot fit the concept of analytical sensitivity. Indeed, it is called "clinical sensitivity" or "diagnostic sensitivity," as correctly indicated later in the same page 1, under "clinical sensitivity."

- **After reviewing the definitions for "analytical sensitivity," the subcommittee decided to remove this term from the document and replace it with the term "minimal detectable concentration (MDC)." The subcommittee agreed that this is more appropriate for the document.**

Section 2

13. "Modifications of the product design, materials, or operating conditions during the course of a clinical evaluation may invalidate the entire study, due to the fact that the analytical performance characteristics of the assay may have changed while the study was in progress."

    Add the following sentence: "Changes should be evaluated for possible impact to the study, and those portions of the study repeated if necessary."

- **The subcommittee believes that consideration of any changes to the assay and the effect on the evaluation are implicit in the paragraph. The text remains unchanged.**

Section 2.1.2

14. Last sentence. "For these assays, the developer must establish a means of ensuring purity of the calibration material and its preservation in a stable matrix, as well as a means of validating its concentration." Change "validating" to "establishing."

- **The text has been revised.**

15. Section 2.1.2: "Calibrators, Standards, and Controls." There is also information on that topic in Sections 2.3.2, 2.4.1, and 2.4.3.

- **Sections 2.3.2, 2.4.1, and 2.4.3 have been revised.**

Section 2.1.4

16. 2nd Paragraph. The statement "source materials were tested for the presence of HIV and hepatitis antigens" should continue to state that they were found negative. Perhaps the testing method should also be mentioned, as it may become relevant for retrospective correlation of aberrant cases.

- **The text has been revised by adding, "and found negative" to the fourth sentence of the second paragraph.**

Section 2.2.1

17. Section 2.2.1: "Type." There is also information regarding interfering substances in Section 2.4.2. Also in Section 2.2.1, would you want to include that the investigator needs to know not only specimen type, but also the minimum amount of specimen necessary to perform the test accurately?

- **Section 2.2.1 and 2.4.2 are not redundant for interfering substances but provide synergistic information. The amount of required specimen is defined by default in the assay instructions, and by the evaluation protocol provided in Section 3.3.**

Section 2.4.1

18. You may wish to combine the documentation of this data in this section rather than in a separate section (now Section 2.5) much further down within the document, or you can create a separate section for ALL documentation required.

- **Section 2.5 has been deleted.**

19. "The analytical performance characteristics of the assay that need to be established may include accuracy, precision ... analytical sensitivity (minimal detectable level), recovery of analyte, and method comparison..." Delete "minimal detectable level" and add "limit of quantitation."

    Rationale: Analytical sensitivity is already defined in the Definitions section and "minimal detectable level" is not complete.

    The limit of quantitation (LOQ) or functional sensitivity is evaluated in immunoassay clinical studies. For example, transplant drug assays often evaluate LOQ based on a predetermined maximum %CV that is acceptable to end users. Similar studies have been done for PSA and TSH.

- **The subcommittee has replaced the term "analytical sensitivity" with the term "minimal detectable concentration (MDC)" throughout the text. In addition, the text in Section 2.4.1 has been revised.**

20. Here the "analytical sensitivity" is not defined as the ratio between response and stimulus (i.e., the slope of the calibration curve), in agreement with the previous definition of Section 1.2 (page 1). Here the same term is defined, in brackets, as the "minimal detectable limit" (i.e., the detection limit), which is not a synonym for "analytical sensitivity" (see EDMA Terminology paper).

- **The term "analytical sensitivity" has been replaced with "minimal detectable concentration (MDC)" and a definition for "minimal detectable concentration" has been added to the Definitions section, as this term is more appropriate.**

21. There is an evident contradiction between the definitions in Sections 1.2 and 2.4.1. Moreover, "accuracy" is not defined, and it is not clear whether it means the combination of "trueness" and "precision," according to ISO/ 5725-1, or only the component of systematic error, called "trueness." This should be clarified, because, in U.S. literature, "accuracy" is too often used synonymously with "trueness," which is not correct, according to ISO standards mentioned above.

- **A definition for "accuracy" has been added to the Definitions section. Moreover, the definitions and terminology throughout the document have been changed to maintain consistency with the ISO standards mentioned above. Please see the "Note on Terminology" in the Foreword and the responses to Comments 1, 22, and 36.**

22. At last, "precision" is here correctly defined as intra-assay, now called "repeatability," and interassay now called "reproducibility." The two new terms (according to VIM: 1993) should appear instead of intra- or interassay, lot-to-lot, or other similar terms.

- **The definitions for "repeatability" and "reproducibility" have been added to the Definitions section and they are used appropriately throughout the text. Descriptions such as lot-to-lot are mentioned in an effort to explain the terms mentioned above. Please see the responses to Comments 1, 21, and 36.**

Section 2.4.3

23. Delete the second and fourth sentences, and revise the third and fifth sentences to read as follows:

    - Samples that read high and outside of the reportable range should be diluted and rerun and those reading lower than the reportable range limit should be reported as below the level of the sensitivity of the assay.

    - In most circumstances, any reading obtained above the analytical sensitivity of the assay or above the minimal detectable concentration is considered within the reportable range of the assay.

    Rationale: The first sentence does not hold true for all assays. The rest of the paragraph is clarification.

- **The section on "Reportable Range" has been deleted, and a definition for "reportable range" has been added to Section 1.2.**

24. What is the difference between analytical sensitivity and MDC? Why mention both? If different, define?

- **Please see the response to Comment 20.**

Section 2.5

25. This doesn't say anything, except that "you should" in the document.

- **The subcommittee agrees and has removed Section 2.5 from the document.**

Section 3.1

26. Safety. The statement should include a requirement for inclusion of detailed instructions for handling/cleaning up spills/accidents. The details should be specific to the infectious risks and reagents involved.

- **This subcommittee believes the current text falls under the scope of good laboratory practice. The subcommittee refers the reader to the current version of NCCLS document M29—*Protection of Laboratory Workers from Occupationally Acquired Infections* for additional information.**

27. Add a section on "Supply inventory control."

    Rationale: Accountability of IUO reagents needs to be added to this section. Governing bodies are concerned that IUO reagents be used for a specific purpose and not diverted and used elsewhere. It is increasingly important to be able to account for all materials that are shipped, used, and destroyed or returned during a clinical evaluation.

- **The text has been revised by adding the bullet, "investigational materials inventory accountability log."**

Section 3.2.1

28. This section should perhaps be worded more strongly regarding the importance of potential risks to the human subjects and the critical nature of that decision. The present language is somewhat weak; however, I qualify my concerns with the fact that I have limited research exposure.

- **The subcommittee appreciates the commenter's concern, but notes that these issues are covered by Institutional Review Board (IRB) review.**

Section 3.2.2

29. Last line. Change "clinical evaluation" to read "portion of clinical evaluation using subjects or subject specimens."

    Rationale: This change in wording allows the site to set up the assay, train, and do initial performance evaluations prior to or concurrent with the IRB review.

- **The subcommittee has revised the text as indicated.**

Section 3.2.5

30. When research results are returned to the subject or a healthcare provider, the facility where the test is performed must have an effective, appropriate CLIA certificate.

- **The subcommittee directs the commenter to Section 3.1, which addresses the need for laboratory certification.**

Section 3.3

31. Page 11, (15) Principal Investigator Responsibilities: The statement should perhaps be a general, inclusive statement of broad responsibility rather than a mere caution.

- **The text has been revised.**

Section 3.4

32. Last two sentences. Be more specific and clear, and give examples.

- **The text has been revised.**

Section 3.5

33. Delete (18).

    Rationale: This seems to be saying the same as item 21.

- **The text has been revised.**

Section 3.6.1

34. First line, "The objectives of a clinical evaluation will depend on the proposed clinical use and indications for use of the immunoassay." Change "proposed clinical" to "intended."

    Rationale: This change puts this sentence more in line with the scope of the document and narrows the requirements of providing clinical use studies for every assay in development, which is sometimes unnecessary and would be burdensome.

- **The first paragraph of Section 3.6.1 has been rewritten.**

35. Third paragraph: Delete the sentence beginning "Use of ROC analysis..."

    Rationale: If one were using cancer diagnostic assays, for example, and only tested patients with cancer, it is unclear how ROC analysis would be effective, since there would be no way to calculate clinical specificity (the x-axis is plotted as 1-specificity).

- **The text has been revised.**

Section 3.6.2

36. In the middle of this chapter there is a sentence, reading: "the analytical performance characteristics, including reproducibility, accuracy, and precision must be...."

    In fact, as stated in Section 2.4.1, precision may be expressed as reproducibility and repeatability. Thus, the sentence should mention precision (i.e., repeatability and reproducibility) or only precision or only repeatability and reproducibility. Otherwise, the reader gets confused, as in this case, where it seems that reproducibility is different from precision, whilst it is one component of precision. Moreover, in this chapter, "accuracy" is again not defined and it is not clear what it means (see comment on Section 2.4.1).

It is absolutely not sufficient to check accuracy and reproducibility (by contrast, it is even wrong) in order to verify the analytical performance of a system.

All the specific performance characteristics have to be checked, including analytical sensitivity, analytical specificity, precision (including repeatability and reproducibility), trueness, and accuracy. In certain cases, also diagnostic (or clinical) sensitivity and specificity have to be checked.

- **The ISO definition for "accuracy" has been added to the Definitions section and it is used appropriately throughout the text. Furthermore, the Definitions section and the terminology have been revised to be consistent with international use. Please see the "Note on Terminology" in the Foreword and the responses to Comments 1, 21, and 22.**

Section 3.8.2

37. Revise the fourth sentence as follows: "Examples of inclusion/exclusion criteria include medication use, presence of a disease state or stage, fasting or dietary requirements, and adequate specimen volume for initial as well as discordant or discrepant retesting."

    Rationale: It is important to allow this type of exclusion within the protocol.

- **The text has been revised to read, "specimen volume and quality" in the fifth sentence.**

Section 3.9

38. Delete the sentence beginning with "Although the use of different test personnel…" and replace it with the following: "Care should be taken by the study sponsor to provide a protocol that allows for elimination of evaluator bias. This may mean that only coded patient samples may be used throughout the protocol."

    Rationale: The requirement to use coded samples in a study using automated instruments is of questionable utility in controlling bias and would add a degree of complexity that may not be necessary.

- **The subcommittee does not agree with the rationale. Not all bias is eliminated by the use of automated instruments.**

Section 4.1

39. This section states the importance of training, but training is also addressed in Section 2.3.4.

- **Section 4.1 has been revised.**

Section 4.4

40. Delete the first sentence and replace it with the following: "Data from clinical evaluations shall be retained in accordance with regulations in countries where the product is being registered or used."

    Rationale: There are various regulations surrounding retention of data which change depending on which country(ies) the product is marketed in. Leave this statement more open.

- **The subcommittee agrees, and the text has been revised.**

Summary of Delegate Comments and Subcommittee Responses

I/LA21-A: *Clinical Evaluation of Immunoassays; Approved Guideline*

General

I saw the proposed version (September 1999) early in the year 2000; and I communicated my comments to various persons of the EDMA Standardization Working Party. My main concerns at that time were that the guideline was definitely NOT in line with current concepts of traceability especially for the Type B analytes, i.e., all the glycoproteins measured by immunoprocedures (cf. prEN ISO 17511). The type B analytes comprise some 600 entries including tumor markers, hormones, virology markers etc.

With the new version, I had hoped to find that in Section 1.2, Section 2 as a whole, Section 3 (particularly 3.6.2 and 3.7.3.) adequate discussion was given to the many problems in this most important area of laboratory medicine. I felt disappointed to find only some vague allusions to them or no discussion at all.

I believe that NCCLS has been fully informed during the years about the new concepts; in my modest opinion, I think they should have been brought to the attention of the subcommittee.

Therefore, I can only come to the conclusion that the I/LA21 guideline is strictly for regional purposes, i.e., within the USA, where many of us work for a global market world. I must confess that it surprises me that so many members of the subcommittee, coming from the U.S.-IVD Industry, have not made objections at an earlier stage of development of this guideline.

As a suggestion, I propose that a number of persons knowledgeable in the new concepts meet with a selection (emphasis on the US IVD Industry) of the members of the subcommittee in order to try to sort out the problems. I have made the same suggestion more than 1.5 years ago, but have never received an appropriate answer.

I am sorry to say that I cannot be more positive about this guideline; perhaps the ideas and the perceptions will change?

- **Please see the response to Comment 1 in the Summary of Comments section regarding the proposed-level version of this guideline.**

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 on comparing the resulting precision estimates with manufacturer precision performance claims and determining the validity of such comparisons; and manufacturer guidelines for establishing claims.

EP6-P2 **Evaluation of the Linearity of Quantitative Analytical Methods; Proposed Guideline (2001).** This document provides guidelines 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.

EP7-P **Interference Testing in Clinical Chemistry; Proposed Guideline (1986).** This document provides background information and procedures for characterizing the effects of interfering substances on test results.

EP9-A **Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline (1995).** This document addresses procedures for determining the bias between two clinical methods or devices; and design of a method comparison experiment using split patient samples and data analysis.

EP10-A **Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline (1998).** This guideline addresses experimental design and data analysis for preliminary performance evaluation of an analytical method or device.

EP14-A **Evaluation of Matrix Effects; Approved Guideline (2000).** This document provides guidance for evaluating the error or bias in analyte measurements that results from the sample matrix (physiological or artificial) when two analytical methods are compared.

GP10-A **Assessment of the Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plots; Approved Guideline (1995).** This document describes the design of a study to evaluate clinical accuracy of laboratory tests; procedures for preparing ROC curves; glossary of terms; and computer software program information.

H3-A4 **Procedures for the Collection of Diagnostic Blood Specimens by Venipuncture; Approved Standard—Fourth Edition (1998).** This document provides procedures for the collection of diagnostic specimens by venipuncture, including line draws, blood culture collection, and venipuncture in children.

I/LA18-A **Specifications for Immunological Testing for Infectious Diseases; Approved Guideline (1994).** This guideline outlines specimen requirements; performance criteria; algorithms for the potential use of sequential or duplicate testing; recommendations for intermethod comparisons of immunological test kits that detect infectious diseases; and specifications for development of reference materials.

*Proposed- and tentative-level documents are being advanced through the NCCLS consensus process; therefore, readers should refer to the most recent editions.

Related NCCLS Publications (Continued)

M29-A2 **Protection of Laboratory Workers from Occupationally Acquired Infections; Approved Guideline – Second Edition (2001).** 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.

NRSCL8-A **Terminology and Definitions for Use in NCCLS Documents; Approved Standard (1998).** This document provides standard definitions for use in NCCLS standards and guidelines, and for submitting candidate reference methods and materials to the National Reference System for the Clinical Laboratory (NRSCL).

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 1-56238-455-4