GP29-A
Vol. 22 No. 26
Replaces GP29-P
Vol. 21 No. 11

Assessment of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline

This document offers methods to assess test performance when proficiency testing (PT) is not available; these methods include examples with statistical analyses. This document is intended for use by laboratory managers and testing personnel in traditional clinical laboratories as well as in point-of-care and bedside testing environments.

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.

Assessment of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline

Abstract

NCCLS document GP29-A—*Assessment of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline* offers methods to assess test performance when formal proficiency testing (PT) programs (also known as external quality assessment (EQA) programs) are not available. The guideline includes examples with statistical analyses. This document is intended for use by laboratory managers and testing personnel in traditional clinical laboratories as well as in point-of-care and bedside testing environments.

NCCLS. *Assessment of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline*. NCCLS document GP29-A (ISBN 1-56238-479-1). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.

GP29-A
ISBN 1-56238-479-1
ISSN 0273-3099

Assessment of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline

Volume 22 Number 26

Stephen J. Sarewitz, M.D., Chairholder
Dorothy J. Ball, Ph.D.
Harvey George, Ph.D.
Glen L. Hortin, M.D.
Stan R. Johnson, M.B.A
Joanne Mei, Ph.D.
W. Gregory Miller, Ph.D.
Albert Rabinovitch, M.D., Ph.D.
Janet M. Smith, B.Sc., M.Sc., FRCPath
Theresa D. Stokeld, MBA, MT(ASCP) DLM
Daniel W. Tholen, M.S.

Suggested Citation

(NCCLS. *Assessment of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline.* NCCLS document GP29-A [ISBN 1-56238-479-1]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.)

Proposed Guideline
July 2001

Approved Guideline
December 2002

ISBN 1-56238-479-1
ISSN 0273-3099

Committee Membership

Area Committee on General Laboratory Practices

Stephen J. Sarewitz, M.D. **Chairholder**	**Valley Medical Center** **Renton, Washington**
Sheila M. Woodcock, MLT, ART, MBA **Vice-Chairholder**	**QSE Consulting** **Nova Scotia, Canada**

Subcommittee on Assessment of Laboratory Tests When Proficiency Testing is Not Available

Stephen J. Sarewitz, M.D. **Chairholder**	**Valley Medical Center** **Renton, Washington**
Dorothy J. Ball, Ph.D.	Abbott Laboratories Irving, Texas
Harvey George, Ph.D.	MA Dept. of Public Health Labs. Jamaica Plain, Massachusetts
Glen L. Hortin, M.D.	National Institutes of Health Bethesda, Maryland
Stan R. Johnson, M.B.A	Beckman Coulter, Inc. Brea, California
Joanne Mei, Ph.D.	Centers for Disease Control and Prevention Atlanta, Georgia
Albert Rabinovitch, M.D., Ph.D.	Specialty Laboratories Santa Monica, California
Janet M. Smith, B.Sc., M.Sc., FRCPath	University Hospital Birmingham Birmingham, United Kingdom
Theresa D. Stokeld, MBA, MT(ASCP) DLM	CHRISOPE Technologies Lake Charles, Louisiana

Advisors

James Berlin	Medical Analysis Systems Miami, Florida
Craig Columbel	Washington Dept. of Health Shoreline, Washington
Joanne Cornbleet, M.D.	Stanford University Medical Center Stanford, California
Thomas R. Fritsche, Ph.D., M.D.	University of Washington Seattle, Washington

Advisors (Continued)

Patricia L. James, M.A.	Agency for Health Care Administration Tallahassee, Florida
James D. MacLowry, M.D.	Oregon Health Sciences University Portland, Oregon
Thomas A. Merrick, M.D.	Presbyterian/St. Luke's Medical Center Denver, Colorado
W. Gregory Miller, Ph.D.	Virginia Commonwealth University Richmond, Virginia
Robert F. Moran, Ph.D., FCCM, FAIC	mvi Sciences Methuen, Massachusetts
Valerie L. Ng, Ph.D., M.D.	San Francisco General Hospital San Francisco, California
John W. Ross, M.D.	Kennestone Hospital Marietta, Georgia
Elkin Simson, M.D., Ch.B., M.Med.	Mount Sinai Medical Center New York, New York
Daniel W. Tholen, M.S.	Statistical Services Traverse City, Michigan
Jennifer K. McGeary, M.T.(ASCP), M.S.H.A. *Staff Liaison*	NCCLS Wayne, Pennsylvania
Patrice E. Polgar *Editor*	NCCLS Wayne, Pennsylvania
Donna M. Wilhelm *Assistant Editor*	NCCLS Wayne, Pennsylvania

Active Membership
(as of 1 October 2002)

Sustaining Members

Abbott Laboratories
American Association for Clinical Chemistry
Beckman Coulter, Inc.
BD and Company
bioMérieux, Inc.
CLMA
College of American Pathologists
GlaxoSmithKline
Nippon Becton Dickinson Co., Ltd.
Ortho-Clinical Diagnostics, Inc.
Pfizer 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.
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 Association of Testing Authorities – Australia
National Society for Histotechnology, Inc.
Ontario Medical Association Quality Management Program-Laboratory Service
RCPA Quality Assurance Programs PTY Limited
Sociedade Brasileira de Analises Clinicas
Sociedad Espanola de Bioquimica Clinica y Patologia Molecular
Taiwanese Committee for Clinical Laboratory Standards (TCCLS)
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
Iowa State Hygienic Laboratory
Massachusetts Department of Public Health Laboratories
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
Alifax S.P.A.
Ammirati Regulatory Consulting
Anaerobe Systems
Asséssor
AstraZeneca
AstraZeneca R & D Boston
Avant Immunotherapeutics, Inc.
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
David G. Rhoads Associates, Inc.
Diagnostics Consultancy
Diagnostic Products Corporation
Eiken Chemical Company, Ltd.
Elan Pharmaceuticals
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.
IGEN 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.)
NimbleGen Systems, 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
Philips Medical Systems
Powers Consulting Services
Premier Inc.
Procter & Gamble Pharmaceuticals, Inc.
The Product Development Group
QSE Consulting
Quintiles, Inc.
Radiometer America, Inc.
Radiometer Medical A/S
Replidyne
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.
Synermed Diagnostic Corp.
Sysmex Corporation (Japan)
Sysmex Corporation (Long Grove, IL)
The Clinical Microbiology Institute
The Toledo Hospital (OH)
Theravance Inc.
Transasia Engineers
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

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)
Allina Health System (MN)

Alton Ochsner Medical Foundation (LA)
Antwerp University Hospital (Belgium)
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)
Brooks Air Force Base (TX)
Broward General Medical Center (FL)
Cadham Provincial Laboratory
Calgary Laboratory Services
Carilion Consolidated Laboratory (VA)
Cathay General Hospital (Taiwan)
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)
Chang Gung Memorial Hospital (Taiwan)
Changi General Hospital (Singapore)
The Charlotte Hungerford Hospital (CT)
Children's Hospital (LA)
Children's Hospital (NE)
Children's Hospital & Clinics (MN)
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 County Hospital (IL)
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)
Diagnosticos da América S/A (Brazil)
Dr. Everett Chalmers Hospital (New Brunswick, Canada)
Doctors Hospital (Bahamas)
Duke University Medical Center (NC)
Dwight David Eisenhower Army Med. Ctr. (GA)
E.A. Conway Medical Center (LA)
Eastern Maine Medical Center
East Side Clinical Laboratory (RI)
Eastern Health (Vic., Australia)
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
Focus Technologies (CA)
Foothills Hospital (Calgary, AB, Canada)
Fresenius Medical Care/Spectra East (NJ)
Fresno Community Hospital and Medical Center
Frye Regional Medical Center (NC)
Gambro BCT (CO)
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)
Hopital du Sacre-Coeur de Montreal (Montreal, Quebec, Canada)
Hôpital Maisonneuve – Rosemont (Montreal, Canada)
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 University
Innova Fairfax Hospital (VA)
Institute of Medical and Veterinary Science (Australia)
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)
L'Hotel-Dieu de Quebec (Canada)
Libero Instituto Univ. Campus BioMedico (Italy)
Louisiana State University Medical Center
Maccabi Medical Care and Health Fund (Israel)
Malcolm Grow USAF Medical Center (MD)
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 Hospital at Gulfport (MS)
Memorial Medical Center (IL)
Memorial Medical Center (LA) Jefferson Davis Hwy
Memorial Medical Center (LA) Napoleon Avenue
Mercy Medical Center (IA)
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)
Nassau County Medical Center (NY)
National Institutes of Health (MD)
Naval Hospital – Corpus Christi (TX)
Naval Surface Warfare Center (IN)
Nebraska Health System
New Britain General Hospital (CT)
New England Fertility Institute (CT)
New Mexico VA Health Care System
New York University Medical Center
North Carolina State Laboratory of Public Health
North Shore – Long Island Jewish Health System Laboratories (NY)
North Shore University Hospital (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)
Pathology Associates Medical Laboratories (WA)
The Permanente Medical Group (CA)
Piedmont Hospital (GA)
Pikeville Methodist Hospital (KY)
Pocono Hospital (PA)
Presbyterian Hospital of Dallas (TX)
Providence Health Care
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's Medical Center (NY)
St. Joseph Mercy Hospital (MI)
St. Jude Children's Research Hospital (TN)
St. Luke's Regional Medical Center (IA)
St. Mary of the Plains Hospital (TX)
St. Mary's Hospital & Medical Center (CO)
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)
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 College Hospital (Galway, Ireland)
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)
Virginia Department of Health
VA (Hines) Medical Center
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)
West Shore Medical Center (MI)
Wilford Hall Medical Center (TX)
William Beaumont Army Medical Center (TX)
William Beaumont Hospital (MI)
William Osler Health Centre (Brampton, ON, Canada)
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

Contents (Continued)

Foreword

Proficiency testing (PT), also known as external quality assessment (EQA), is an important part of quality management in the clinical laboratory. PT complements internal quality control to help assure that patient test results are valid. In many cases, governmental agencies that oversee clinical laboratories require participation in a formal PT program.

However, formal PT programs are not available for a substantial number of laboratory tests. The reasons vary. Some analytes are unstable, precluding the preparation of PT materials, or matrix effects may prevent reliable analysis. Some tests are performed in only a few laboratories, so that it is not practical to develop a formal PT program. PT is not available for certain pathogenic microorganisms because of the hazards of transporting the organisms.

This document offers methods to assess test performance when PT is not available. These methods are termed "alternative assessment procedures," or AAPs.[a] The document addresses a variety of tests, including quantitative analyses of blood, microbiological cultures, morphologic analyses, and *in vivo* tests. The options for some of these tests are necessarily rather limited.

The document may be useful for managers, supervisors, and laboratory personnel in traditional laboratories, as well as personnel performing point-of-care, clinic, and bedside testing.

Key Words

Alternative assessment procedure, external quality assessment, proficiency testing

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 deserves 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 guideline, it is necessary to point out that several terms are used differently in the USA and other countries, notably those in Europe.

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 an accepted reference value. 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.

[a] Neither PT nor the procedures described in this document are adequate *by themselves* to comprehensively validate a test method.

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:

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

GP29-A 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			

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

Assessment of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline

1 Introduction

Proficiency testing (PT), also known as external quality assessment (EQA), is a valuable element in clinical laboratory quality management.[1] Many agencies and associations, governmental and non-governmental, offer PT for numerous analytes. However, for a variety of reasons, PT is not available for a substantial number of laboratory tests. This document offers guidance to clinical laboratories in the development of alternative quality assessment procedures (AAPs) when PT is not available.

2 Scope

The guidelines in this document apply to clinical laboratory tests performed in the traditional laboratory setting, as well as point-of-care testing, and testing in clinics and physician offices. The scope of the document does not include home testing (i.e., patient self-testing). Quality assessment programs for home testing have been described.[2]

This document makes no distinction between regulated and nonregulated analytes, as defined in the U.S. under the Clinical Laboratory Improvement Amendments of 1988 (CLIA '88).[3] The subcommittee believes that assessment procedures (either PT or AAPs) to validate ongoing performance are important for all laboratory tests.

This document suggests general approaches and provides examples, but does not prescribe specific assessment procedures for individual analytes. The responsibility for selecting specific assessment procedures lies with the individual clinical laboratory.

3 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*.

4 Definitions and Abbreviations[b]

4.1 Definitions

Accepted reference value – A value that serves as an agreed-upon reference for comparison and which is derived as: a theoretical or established value based on scientific principles; an assigned value based on

[b] 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.

experimental work of some national or international organization; or a consensus value based on collaborative experimental work under the auspices of a scientific or engineering group.

Accuracy – Closeness of the agreement between the result of a measurement and an accepted reference value of the measurand/analyte; **NOTE**: See Note on Terminology in the Foreword.

Assessment event – A test assessment procedure(s) performed at one point in time.

Audit-sample testing – Testing of stored aliquots from a biologic specimen repeatedly over time in a specific assay system.

Average: See **Mean**.

Bias – 1) The systematic, signed deviation of the test results from the accepted reference value; **NOTES:** a) Defined in ***(ISO3534-1/93-3.13)*** as "the difference between the expectation of the test results and an accepted reference value"; b) In general, the deviation/difference is based on replicate measurement using an accepted reference comparison method and the method being tested, and expressed in the units of the measurement or as a percentage; **2) Interinstrument/intermethod or interlaboratory bias**, *n* - The difference observed by comparing two specified instruments' or laboratories' methods under specified conditions of analysis, concentration range, method, etc.

Calibration material//Calibrator – A material or device of known, or assigned quantitative characteristics (e.g., concentration, activity, intensity, reactivity, responsiveness) used to adjust the output of a measurement procedure or to compare the response obtained with the response of a test specimen and/or sample.

Coefficient of variation (CV) – For a non-negative characteristic, the ratio of the standard deviation to the average; **NOTE:** It is often multiplied by 100 and expressed as a percentage.

Common cause variation – Variation resulting from sources inherent in the testing process; **NOTE:** Also known as "random variation" or "process variation."

Commutable – Degree to which a material yields the same numerical relationships between results of measurements by a given set of measurement procedures, purporting to measure the same quantity, as those between the expectations of the relationships for the same procedures applied to those types of material for which the procedures are intended; **NOTE:** The degree to which a given material demonstrates intermethod changes in response (for two or more measurement procedures) comparable to the changes in response observed for relevant materials, is an indication of the commutability of the material among the specified measurement procedures.

External quality assessment: See **Proficiency testing**.

Imprecision: See **Precision**.

Manufacturer's product calibrator – The material or materials provided by a manufacturer for the purpose of establishing the measurement relationships of an *in vitro* diagnostic device.

Matrix effect – The influence of a sample property, other than the measurand/analyte, on the measurement, and thereby on the value of the measurand.

Mean – The sum of values divided by the number of values; **NOTE:** The term "average" [is used] when referring to the result of a calculation on the data obtained in a sample.

Power of error detection – The statistical probability that a quality control system will detect changes that exceed defined limits.

Precision – The closeness of agreement between independent test results obtained under prescribed/stipulated conditions; **NOTE:** Precision is typically expressed quantitatively in terms of imprecision—the SD or the CV of the results in a set of replicate measurements.

Proficiency testing//External quality assessment (PT/EQA) – Determination of laboratory testing performance by means of interlaboratory comparisons; **NOTES:** a) Commonly, a program periodically sends multiple specimens to members of a group of laboratories for analysis and/or identification; the program then compares each laboratory's results with those of other laboratories in the group and/or with an assigned value, and reports the results to the participating laboratory and others; b) Other forms of PT/EQA include: data transformation exercises, single-item testing (where one item is sent to a number of laboratories sequentially and returned to the program at intervals), and one-off exercises (where laboratories are provided with a test item on a single occasion).

Quality – **1)** Degree to which a set of inherent characteristics fulfills requirements; **NOTES:** a) The term "quality" can be used with adjectives such as poor, good, or excellent; b) "Inherent," as opposed to "assigned," means existing in something, especially as a permanent characteristic; **2)** Doing the right thing, right, in a timely manner.

Quality control (QC) – The operational techniques and activities that are used to fulfill requirements for quality.

Quality management – Coordinated activities to direct and control an organization with regard to quality; **NOTE:** Direction and control with regard to quality generally includes establishment of the quality policy and quality objectives, quality planning, quality control, quality assurance, and quality improvement.

Reference method – A thoroughly investigated method, in which exact and clear descriptions of the necessary conditions and procedures are given for the accurate determination of one or more property values, and in which documented trueness and precision of the method are commensurate with the method's use for assessing the trueness of other methods for measuring the same property values, or for assigning reference method values to reference materials.

Regression analysis – **1)** The process of estimating the parameters of a model by optimizing the value of an objective function and then testing the resulting predictions for statistical significance against an appropriate null hypothesis model; **2)** The process of describing mathematically the relationship between two or more variables; **NOTE:** This can include the parametric testing of the statistical significance of the relationship, if random errors are assumed to be normal; **3) Linear regression** – a) A statistical calculation that results in parameters that describe the assumed linear relationship between values of an independent and a dependent variable wherein the independent variable is known exactly; b) The process of estimating the straight-line relationship between an independent and dependent variable, where the independent variable is known without error; **4) Orthogonal regression** – a) A statistical calculation that results in parameters that describe the assumed linear relationship between values of an independent and a dependent variable wherein the independent variable is assumed to vary normally; b) The process of estimating the straight-line relationship between an independent and dependent variable, where the independent variable is not known exactly.

Sensitivity – **1)** *In quantitative testing*, the change in response of a measuring {system or} instrument divided by the corresponding change in the stimulus; **2)** *In the context of QC*, the power of error detection of a QC system.

Specificity – 1) The ability of a test or procedure to correctly identify or quantify an entity in the presence of interfering phenomena/influence quantities; **2)** *In the context of QC*, the probability that a QC system will indicate absence of special cause variation (i.e., process error) when special cause variation is truly absent; 1 minus the probability of "false alarms" wherein QC data points exceed tolerance limits yet no error can be identified in the test system.

Special cause variation – Variation from sources outside the testing process; also known as "assignable cause variation," or "process error"; **NOTE:** Sources of special cause variation include interferences, operator error, instrument malfunction, and deterioration of reagents.

Split-sample testing – A single biologic sample (e.g., tube of blood) divided into aliquots, wherein one aliquot is tested on a particular assay system (or by a particular analyst), other aliquot(s) are tested on other instruments(s) (or by other analysts[s]) (often in a different laboratory), and the results are compared.

Split-specimen testing – A test involving two biologic specimens obtained at the same time from the same source on two or more different assay systems (or by two or more analysts), often in two or more different laboratories; **NOTE:** For example, testing two tubes of blood collected from a particular patient during the same venipuncture.

[Sample] standard deviation (SD) – **1)** A measure of variability//dispersion that is the positive square root of the population variance; **2)** The quantity characterizing the dispersion of the results for a series of measurements of the same measurand.

Trueness – The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value; **NOTES:** a) See also **Accuracy, Bias;** b) See Note on Terminology in the Foreword.

Trueness control material – Reference material that is used to assess the bias of measurement of a measuring system.

Validation – Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled. [ISO 9000:2000; ISO/FDIS 17511]

Variance – A measure of dispersion of observations in a sample/population which is the sum of the squared deviations of observations from their average divided by {the degrees of freedom}; **NOTE**: Usually, the degrees of freedom equals one less than the number of observations.

4.2 Abbreviations

Abbreviations used in this document:

AAP	Alternative assessment procedure
CLIA	Clinical Laboratory Improvement Amendments (U.S.A.)
CV	Coefficient of variation
EQA	External quality assessment
PT	Proficiency testing
QC	Quality control
SD	Standard deviation

5 Rationale

Clinical laboratories use internal quality control (QC) procedures as a primary tool to assure the validity of patient test data. For quantitative assays, these procedures generally employ manufactured materials that are tested along with patient specimens. Routine QC allows laboratorians to separate variation that is inherent in the testing process—i.e., common cause variation—from *special cause variation* resulting from an abnormal condition affecting the testing process, such as operator error, reagent problems, incorrect calibration, or instrument malfunction. However, QC has limitations. Among them are the following:

- QC is not perfectly sensitive or specific; it does not detect all instances of special cause variation, and it sometimes inappropriately flags variation inherent in the testing process (i.e., false rejection).
- QC does not necessarily assess test trueness.
- QC does not assess comparability of results with other laboratories.

PT/EQA can serve as an additional quality monitor to address these limitations:

- PT may detect problems/errors that were not detected by the internal QC system. See the most current edition of NCCLS document GP27—*Using Proficiency Testing (PT) to Improve the Clinical Laboratory*.
- When analytes in PT materials can be traced to a reference method, a laboratory can determine the accuracy of its analysis under some circumstances (e.g., in the absence of significant matrix effects).[4]
- Participation in a PT program enables a laboratory to compare its performance against other laboratories using similar methods/reagents/instruments.

The functions of PT/EQA have been described in the literature.[1,5-12] (See also the most current edition of NCCLS document GP27— *Using Proficiency Testing (PT) to Improve the Clinical Laboratory*.)

PT/EQA is, however, not available for many tests. For such tests, laboratories should, when appropriate and practical, implement an alternative assessment procedure (AAP[c]).

Certain governmental and nongovernmental certifying and accrediting bodies require participation in PT and also require laboratories to implement AAPs in the absence of PT.[d] However, the subcommittee believes that AAPs are important quality elements in their own right, regardless of the requirements of certifying/accrediting bodies.

Laboratories (including those performing unique or low-volume analyses—e.g., research laboratories) should try to develop AAPs that provide information similar to that provided by participation in PT. For example, patient specimens may be sent to another laboratory(ies) so that data on interlaboratory comparability is generated (e.g., split-sample procedures; see below). If an AAP can be traced to a

[c] Neither proficiency testing nor the alternative assessment procedures discussed in this document are adequate, in themselves, to validate a test method.

[d] In the U.S.A. under the Clinical Laboratory Improvement Amendments of 1988 (CLIA '88), all laboratories reporting patient results from nonwaived tests must participate in PT for certain analytes ("regulated analytes") and implement AAPs for other tests.[13, 14] The College of American Pathologists,[15] ISO,[16, 17] Clinical Pathology Accreditation (UK), [18] and Joint Commission on Accreditation of Healthcare Organizations[19] also require implementing AAPs when PT is not available. ISO document *Quality Management in the Medical Laboratory* requires that laboratories develop AAPs in the absence of PT.[17]

reference method, accuracy can be assessed. Even if neither interlaboratory comparison nor evaluation of accuracy is practical for a particular test, it is still worthwhile to use an AAP to complement QC, because QC is not perfectly sensitive or specific.

AAPs may often use patient specimens, which have certain advantages over the manufactured materials frequently used in PT.

- Matrix effects are reduced when patient specimens are used.

- The steps in the preanalytic phase of clinical patient testing—specimen acquisition, transportation, and processing—are not evaluated by PT programs using manufactured testing materials, because the preanalytic phase of PT differs from patient testing.[e] In contrast, variables related to preanalytic processing may be evaluated by an AAP that uses patient specimens. Patient specimens used for AAPs require attention to stability during storage and transportation between laboratories, to minimize introduction of additional variability not related to clinical testing performance.

In-house AAPs provide more timely data than does a PT program.

6 Tests for Which PT is Not Available[f]

Tests for which PT may not be available include, but are not limited to:

- Newly developed tests

- Uncommonly performed/esoteric tests:

  – antibodies to certain organisms (i.e., *Bordetella pertussis*, *Histoplasma*, *Blastomyces*, *Influenza A* and *B*, *Parvovirus*, *Legionella*);
  – skeletal muscle antibodies;
  – pancreatic polypeptide;
  – CSF myelin basic protein;
  – lipoprotein (a);
  – whole-blood lactate;
  – vitamin A; and
  – β-carotene.

- Certain drugs:

  – felbamate; and
  – gabapentin.

- Tests associated with PT material problems:

  – instability of material or lability of analyte (e.g., erythrocyte osmotic fragility, erythrocyte sucrose hemolysis, cold agglutinins, serum acetoacetate, isoenzymes, serum ammonia, cryoglobulins, stool leukocyte counts, nasal smear eosinophil count, breath tests);
  – cellular function assays (e.g., platelet aggregation studies, studies of neutrophil or lymphocyte function, semen analysis);

[e] Most laboratory errors occur in the preanalytic and postanalytic (results reporting) phases.[1,20,21] The critical domain of quality management of specimen identification, accessioning, and results reporting lies outside the scope of this document.

[f] To the knowledge of the subcommittee, PT was not available for the analytes/tests listed in this section at the time of writing.

  - matrix effects (e.g., free-drug assays, free-hormone assays (testosterone, insulin), free prostate-specific antigen);
  - contamination in highly sensitive analyses (e.g., molecular amplification techniques); and
  - inability of vendor to provide sufficient material for market demand (e.g., hemoglobin abnormalities, tumor markers, whole-blood cytogenetics).

- Tests associated with container-analyte interactions:

  - drug assays;
  - free-hormone assays; and
  - trace metal assays.

- Tests requiring extensive manipulation of sample, such as detection of markers of environmental exposure or injury:

  - chemical and biologic toxins;
  - toxin metabolites (i.e., breakdown products of toxins);
  - protein and DNA adducts; and
  - heavy metals.

- Analytes in unusual matrices/milieus:

  - interstitial fluid (glucose);
  - stool (cholesterol, yeast culture, leukocyte counts);
  - saliva (therapeutic drug monitoring, detection of drugs of abuse, ethanol, serology, hormones);
  - hair analysis (detection of drugs of abuse);
  - dried blood spots (detection of drugs of abuse; therapeutic drug monitoring); and
  - whole blood (HDL).

- Microbiology issues:

  - fastidious, difficult-to-grow microorganisms (i.e., *Helicobacter pylori*);
  - minimal inhibitory antibiotic concentrations for anaerobes;
  - serotyping when number of serotypes is large (i.e., *Salmonella* species);
  - DNA fingerprinting—number of strains of organisms is large; and
  - hazardous organisms (biosafety levels 3 and 4) (i.e., dimorphic fungi [*Coccidioides, Histoplasma* species], *Salmonella typhi*, *Yersinia pestis)*.

- *In vivo* testing:

  - bleeding time;
  - sweat test collection procedure;
  - reflectance bilirubinometry;
  - near-IF whole-blood glucose;
  - breath tests (alcohol, urea, hydrogen, end-tidal CO_2, acetone);
  - indwelling arterial blood gas monitoring;
  - pulse oximetry;
  - monitoring of anesthetic gas concentrations;
  - Schilling test; and
  - fetal scalp pH.

- Geography: laboratory located in area where relevant PT is not offered.

7 Alternative Assessment Procedures

The laboratory should identify those tests for which PT/EQA is not available and develop an AAP for such tests whenever possible. AAPs should be documented in the laboratory procedure manual. Each laboratory should define the frequency of performance and procedures for evaluation of results. In many cases, performing an AAP twice yearly is adequate.

The laboratory should define the limits of acceptability for each quantitative assessment procedure *in advance,* prior to running the procedure. Laboratories may develop limits of acceptability from internal quality control data (e.g., +/- 2 or 3 standard deviations from the mean), provided that sufficient QC data exist; or from data in the literature—i.e., standard-based limits derived from biologic variation or clinical decision points.[22-24] A procedure for developing analytic bias and imprecision (uncertainty) tolerance limits from patient data has been described,[25, 26] although a large patient database is required (20,000 test values). A summary of statistical methods for evaluation of PT data is available.[27] This information may be helpful to laboratories in analyzing results of AAPs.

Over time (i.e., multiple assessment events over years), alternative assessment procedures should employ samples across the clinically relevant range of the analysis.

Results of AAPs should be documented and retained by the laboratory so that trends can be identified. Corrective action in response to unacceptable results should be documented.

Some AAPs use patient samples/data. As noted above, advantages of using patient results include independence from the routine QC system; avoidance of matrix effects; and the capability of evaluating preanalytic factors, such as the effect of collection systems (e.g., gel-containing blood collection tubes[28]), quality of phlebotomy procedure, delays in processing, etc. In addition, the external split-sample procedure (described below) provides interlaboratory comparison. When adopting split-sample procedures, laboratories should consider their institutions' requirements for informed patient consent and maintenance of patient confidentiality.

7.1 Split-Sample Procedures[21, 29]

7.1.1 Split Sample with Another Laboratory

One common procedure for externally verifying test results is to send aliquots of samples[g] to another laboratory(ies) for testing. Split-sample procedures evaluate interlaboratory agreement and testing errors, but do not evaluate trueness (i.e., bias) *per se* unless the outside laboratory uses a method that is calibrated to a reference method or reference material.[h] It is the responsibility of the individual laboratory to determine the appropriate number of samples/specimens to send for split-sample testing. For many analytes, two samples/specimens per assessment event is adequate.

Investigators at the Centers for Disease Control (U.S.A.) studied the power of split-sample testing to detect problems in the analysis of serum total cholesterol and potassium. In this study, the absence of a discrepancy between samples in the split-sample procedure strongly predicted that the original result was correct (negative predictive value of 93 to 100%); however, the presence of discrepancy in the split-sample procedure was less predictive of error in the original result (positive predictive value of 43 to 67%).[21]

[g] In *split-sample* testing, a single biologic sample (e.g., tube of blood) is divided into aliquots. If testing is performed on different biologic samples obtained at the same time (e.g., two tubes of blood obtained at the same venipuncture) then the term *split-specimen* testing is used. These procedures are treated identically for the purposes of this document.

[h] Evaluation of trueness may be problematic even in organized PT programs.[4]

Please refer to Appendixes A and B for an example of a procedure to determine limits of acceptability for results of quantitative split-sample procedures.

7.1.2 Internal Split-Sample Procedures

Examples of internal split-sample procedures include:

(1) Rerun patient sample by a different method; and

(2) For operator-dependent testing, rerun test using a different operator (for example, in morphologic analysis).

7.2 Audit-Sample Procedure

Aliquots of a patient specimen are stored by the laboratory and analyzed periodically across time.[i] Periodic analysis of aliquots of the audit sample assesses reproducibility and stability of calibration of the assay. The audit-sample procedure does not evaluate accuracy (i.e., bias),[21] nor does it provide interlaboratory comparison.

7.3 Analysis of Manufacturer's Product Calibrator or Trueness Control Material

The calibration material provided by the method manufacturer, or another reference material which has been documented to be either commutable with patients' specimens for the test procedure, or traceable to a reference material or procedure,[30] can be used to confirm correct performance of a method. When the method manufacturer's calibration material or trueness control material is used for an AAP, it is best to use a different lot of material from that used for method calibration. Caution should be used, however, because a lot of calibration material may be specific for a lot of reagent. (Note: It is suggested to use manufacturers' product calibrator or trueness control material only when there is no other alternative material or process to provide validation of method performance.)

7.4 Analysis of Interlaboratory Quality Control Data

This assessment procedure includes participation in peer comparison programs that evaluate quality control data submitted from multiple laboratories. Many manufacturers have such programs. Usually, however, when PT is not available for a particular analyte, neither are peer comparison programs.

7.5 Analysis of Patient Data

7.5.1 Averages of Patient Data

There is extensive literature describing the use of patient data for the quality control of clinical laboratory measurements. In the 1950s and 1960s, tracking the daily average of hematology measurements such as hemoglobin, hematocrit, and red cell counts was proposed as a quality control measure.[31,32] This monitoring of the averages of patient data came into widespread application in the 1970s in a form commonly known as "Bull's algorithm."[33-36] This technique compared the average of 20 consecutive patient values versus an established patient mean value. Methods to monitor a daily mean or an "average of normals" need not be limited to hematology and have been applied as quality control measures for a wide variety of clinical laboratory tests.[21, 23,28,31-44] (See also the most current edition of NCCLS document H38—*Calibration and Quality Control of Automated Hematology Analyzers*.)

[i] After an aliquot is removed from storage and tested, it should be discarded and not returned to storage. Returning an aliquot to storage for later testing may compromise specimen integrity because of denaturation, evaporation, contamination, etc.

The assumption made for this approach is the average result of a group of specimens will be relatively constant when the test procedure is stable. For this condition to be true, the results included in the mean must not include values which are outliers relative to the reference population distribution. This technique is best applied for test procedures which have a fairly high volume of results during a reasonably short time period. However, the approach can be considered for lower-volume tests when a population of test specimens can be identified that are expected to fall in a predictable distribution of results. In an acute care/hospital setting, if the laboratory can identify specific days when it receives an increased proportion of abnormal specimens (e.g., weekends, or days when specimens are received from an oncology clinic or dialysis facility), the best practice might be to exclude the patient data obtained on those days from the calculations. The literature cited above provides guidance on selecting suitable numbers of observations and acceptance criteria.

7.5.2 Reference Ranges

Usually, reference ranges are used within a laboratory to provide information for evaluation of individual patient data. Here we are proposing that periodic reevaluation of reference ranges can be used to validate test procedure stability within a laboratory and to verify agreement between laboratories.[45-49] For this approach to be effective, the original reference range determination must be robust and clinically appropriate for the population served by the laboratory, and the new sample must represent the same reference population with the same preanalytical parameters. The general approach based on NCCLS document C28—*How to Define and Determine Reference Intervals in the Clinical Laboratory* is to obtain test procedure results for a minimum of 20 subjects. A nonparametric analysis that 18 of the 20 results are within the original reference range validates the continuing applicability of that range with an approximate 7% probability of false rejection of the existing reference interval. If the criteria are not met, a second 20 specimens should be obtained and the evaluation repeated. Failure to validate the reference range would initiate a more detailed investigation to determine if the cause was due to problems with the analytic test procedure, with preanalytic conditions for specimen collection and handling, or with sampling the appropriate healthy population.

When larger numbers of results are available (for example, from a computer query over a time period of weeks or months), histograms of the result distributions can be prepared and compared to previous time periods and/or to other laboratories' results. Considerations of outlier identification and exclusion similar to that described for the "average of normals" technique can be applied to obtain more homogeneous populations for purposes of comparing stable population groups. Several statistical approaches have been described to extract adequate reference range data from hospitalized and clinic patient populations that can be used for AAPs.[23,40-42,45-50]

7.5.3 Delta-checking

Delta-checking (i.e., evaluating changes in analyte results from an individual patient across time) is usually used to identify an individual patient's result that is suspect because of deviation from previous results for the same patient. Although delta-checking could be used as an AAP, it is usually considered to be part of routine QC.[51] When delta-checking is used as an AAP it may be difficult to determine if an observed change is due to a clinical change in the patient's condition or to a test procedure malfunction.

7.6 Reevaluation of Morphologic Analyses

Reevaluation of morphologic analyses includes:

- review of glass slides by supervisory personnel; and
- review of "unknown" glass slide sets.

7.7 Direct Observation of Technique-Dependent Tests

Personnel performing technique-dependent tests (e.g., sweat test, bleeding time) may be observed by experienced senior analysts or supervisors. A checklist delineating the factors to be observed should be used in the evaluation.

7.8 Clinical Correlation Studies

Clinical correlation studies have limited application to routine test assessment, because of the imperfect correlation of clinical events to laboratory results, as well as biases that may be operating (e.g., test referral bias, disease classification bias). However, correlation studies may be useful in certain circumstances, when a specific disorder can be diagnosed or strongly suggested if the laboratory test result exceeds a threshold value, and the presence of the disorder can be independently determined at a reasonable point in time after testing. Examples include serum CK-MB or troponin in myocardial infarction, or amylase in acute pancreatitis.

7.9 Approaches for *In Vivo* Breath Tests

Hydrogen breath test analyzers are calibrated from a gas cylinder supplied by the manufacturer that contains a stated amount of hydrogen (usually 94 to 96 ppm). To validate the calibration, the alternatives are limited but include 1) obtaining a hydrogen gas cylinder from another manufacturer that contains gas at a different concentration; and 2) using a reference method such as gas-liquid chromatography. Sampling technique can be validated by obtaining and running duplicate patient samples.[j]

7.10 Surrogate Organisms

Culture of attenuated strains or morphologically similar organisms can be used as an AAP for culture of dangerous organisms.

7.11 Use of a PT/EQA Provider from Another Country/Region

A particular analyte may be offered by a PT provider that does not routinely service the laboratory's area. However, timely transport of PT specimens across international borders may be difficult to achieve.

7.12 Government and University Interlaboratory Comparison Programs

For some population-based testing that has high sample throughput and an important public health function, but only a small number of laboratories conducting testing, government or university reference laboratories provide interlaboratory comparison programs. Examples include testing for very long chain fatty acids, dried blood spot assays for inborn errors of metabolism in neonates, and genetic testing.

8 Analysis of Data from Qualitative Alternative Assessment Procedures

Results of tests that have binary results—e.g., positive or negative—can be easily validated if there are ways to determine definitive diagnosis (refer to the most current edition of NCCLS document EP12—*User Protocol for Evaluation of Qualitative Test Performance)*. When there is no definitive diagnosis, or comparisons are needed between methods or laboratories, split-sample testing can be valuable, but it poses a particular problem because of the likelihood of chance agreement. Please refer to Appendix C for a statistical technique to evaluate data from split-sample studies of qualitative tests. This situation is also discussed in NCCLS document EP12—*User Protocol for Evaluation of Qualitative Test Performance*.

[j] Personal communication with David Andrews, Ph.D., Clinical Biochemistry, University Hospital Birmingham NHS Trust, Birmingham, UK.

References

[1] Shahangian S. Proficiency testing in laboratory medicine. *Arch Path Lab Med*. 1998;122:15-30.

[2] Schrot RJ, Foulis PR, Morrison AD, Farese RV. A computerized model for home glucose monitoring proficiency testing: efficacy of an innovative testing program. *Diabetes Educator*. 1999;25(1):48-55.

[3] Department of Health and Human Services, Health Care Financing Administration. Clinical laboratory improvement amendments of 1988; final rule. *Federal Register*. 1992(Feb 28):7139-7186 [42CFR493.801-494.51].

[4] Ross JW, Miller WG, Myers GL, Praestgaard J. The accuracy of laboratory measurements in clinical chemistry. *Arch Pathol Lab Med.* 1998;122:587-608.

[5] Isenberg HD, D'Amato RF. Does proficiency testing meet its objective? *J Clin Micro.* 1996;34(11):2643-2644.

[6] ASTM E1301-95. Standard Guide for Proficiency Testing by Interlaboratory Comparisons. West Conshohocken, PA: ASTM;1995.

[7] Standards for an interlaboratory (proficiency) testing program. *Am J Clin Path*. 1976;66:276-278.

[8] ISO/TEC Guide 43-1. Proficiency Testing by Interlaboratory Comparisons. Part 1. Development and Operation of Proficiency Testing Schemes. Geneva: International Organization for Standardization; 1997.

[9] Ehrmeyer SS, Laessig RH, Schell K. Use of alternative rules (other than the 1_{2s}) for evaluating interlaboratory performance data. *Clin Chem*. 1988;34(2):250-256.

[10] Steindel SJ, Howanitz PJ, Renner SW. Reasons for proficiency testing failures in clinical chemistry and blood gas analysis. *Arch Pathol Lab Med*. 1996;120:1094-1101.

[11] Laessig RH, Ehrmeyer SS, Lanphear BJ, et al. Limits of proficiency testing under CLIA '67. *Clin Chem*. 1992;38(7):1237-1244.

[12] Thompson M, Wood R. International Harmonized Protocol for Proficiency Testing of (Chemical) Analytical Laboratories. *Journal of AOAC International*. 1993;76(4):926-940.

[13] Department of Health and Human Services, Health Care Financing Administration. Clinical laboratory improvement amendments of 1988; final rule. *Federal Register* 1992(Feb28):7146 [42CFR493.801].

[14] Department of Health and Human Services, Health Care Financing Administration. Medicare, and CLIA programs; CLIA fee collection; correction and final rule. *Federal Register*. 1993(Jan 19):5232 [42CFR493.1709].

[15] Commission on Laboratory Accreditation, College of American Pathologists. Inspection Checklists. Northfield, IL: CAP; 2001.

[16] ISO/IEC 17025. ISO/IEC. International Standard. General requirements for the competence of testing and calibration laboratories. 1999;5.9:19.

[17] prEN ISO/IEC 15189. Quality management in the clinical laboratory. Nov 1999;5.6.5.

[18] Standards for the Medical Laboratory. H5.1. January 2001. Clinical Pathology Accreditation (UK) Ltd. www.cpa-uk.co.uk.

[19] 1998-1999 Comprehensive Accreditation Manual for Pathology and Clinical Services. Joint Commission on Accreditation of Healthcare Organizations. Standard QC 1.1.4:297.

[20] Lapworth R, Teal TK. Laboratory blunders revisited. *Ann Clin Biochem*. 1994;31:78-84.

[21] Cembrowski GS. Thoughts on quality-control systems: a laboratorian's perspective. *Clin Chem*. 1997;43(5):886-892.

[22] Shahangian S, Cohn RD, Gaunt EE, Krolak JM. System to monitor a portion of the total testing process in medical clinics and laboratories: evaluation of a split-specimen design. *Clin Chem.*1999;45(2):269-280.

[23] Cembrowski GS. The pursuit of quality in clinical laboratory analyses. *Clin Chem*. 1990; 36(9):1602-1604.

[24] Westgard JO, Seehafer JJ, Barry PL. Allowable imprecision for laboratory tests based on clinical and analytical test outcome criteria. *Clin Chem*. 1994;40:1909-1914.

[25] Klee G. A conceptual model for establishing tolerance limits for analytic bias and imprecision based on variations in population test distributions. *Clin Chim Acta*. 1997;260:175-188.

[26] Klee GG, Schryver PG, Kisabeth RM. Analytic bias specifications based on the analysis of effects on performance of medical guidelines. *Scand J Clin Lab Invest*. 1999;59:509-512.

[27] ISO/IEC Guide 43-1. Proficiency testing by interlaboratory comparisons—Part 1. Development and operation of proficiency testing schemes. 2nd ed.; 1997.

[28] Douville PD, Cembrowski GS Strauss JF. Evaluation of the average of patients: application to endocrine assays. *Clin Chim Acta*. 1987;167:173-185.

[29] Shahangian S, Cohn RD. Variability of laboratory test results. *Am J Clin Pathol*. 2000;113:521-527.

[30] International Organization for Standardization. *In vitro diagnostic medical devices - Measurement of quantities in biological samples - Metrological traceability of values assigned to calibrators and control materials*. ISO/FDIS 17511. Geneva: 2001.

[31] Dorsey DB. Quality control in hematology. *Am J Clin Pathol*. 1963;40:457-464.

[32] Waid ME, Hoffmann RG. The quality control of laboratory precision. *Am J Clin Pathol.* 1955; 25:585-594.

[33] Cembrowski GS. Use of patient data for quality control. *Clin Lab Med*. 1986;6(4):715-733.

[34] Cembrowski GS, Lunetzky ES, Patrick CC, Wilson MK. An optimized quality control procedure for hematology analyzers with the use of retained patient specimens. *Am J Clin Pathol*. 1988;89:203-210.

[35] Bull BS, Elashoff RM, Heilbron DC, et al. A study of various estimators for the deviation of quality control procedures from patient erythrocyte indices. *Am J Clin Pathol*. 1974;61:473-481.

[36] Cembrowski GS, Westgard JO. Quality control of multichannel hematology analyzers: evaluation of Bull's algorithm. *Am J Clin Pathol*. 1985;83:337-345.

[37] Dixon K, Northam BE. Quality control using the daily mean. *Clin Chim Acta.* 1970;30:453-461.

[38] Begtrup H, Leroy S, Thyregod P, Wallöe-Hansen P. 'Average of normals' used as control of accuracy and a comparison with other controls. *Scand J Clin Lab Invest*. 1971;27:247-253.

[39] Talamo TS, Losos FJ, Gebhardt WD, Kessler GF. Microcomputer assisted hematology quality control using a modified average of normals program. *Am J Clin Pathol*. 1981;76:707-712.

[40] Cembrowski GS, Chandler EP, Westgard JO. Assessment of "average of normals" quality control procedures and guidelines for implementation. *Am J Clin Pathol*. 1984;81:492-499.

[41] Woo J, LeFever D, Winkelman JW. Use of "average of normals" quality control procedure in the detection and resolution of assay discrepancies. *Am J Clin Pathol*. 1988;89:125-129.

[42] Westgard JO, Smith FA, Mountain PJ, Boss S. Design and assessment of average of normals (AON) patient data algorithms to maximize run lengths for automatic process control. *Clin Chem*. 1996;42:1683-1688.

[43] Parvin CA, Gronowski AM. Effect of analytical run length on quality control (QC) performance and the QC planning process. *Clin Chem*. 1997;43:2149-2154.

[44] Ye JJ, Ingels SC, Parvin CA. Performance evaluation and planning for patient-based quality control procedures. *Am J Clin Pathol*. 2000;113:240-248.

[45] Amador E, Hsi BP. Indirect methods for estimating the normal range. *Am J Clin Pathol*. 1969;52:538-546.

[46] Harwood SJ, Cole GW. Reference values based on hospital admission laboratory data. *JAMA*. 1978;240:270-274.

[47] Baadenhuijsen H, Smit JC. Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure. *J Clin Chem Clin Biochem*. 1985;23:829-839.

[48] Swaanenburg JCJM, Rutten WPF, Holdrinet ACJM, van Strik R. The determination of reference values for hematologic parameters using results obtained from patient populations. *Am J Clin Pathol*. 1987;88:182-191.

[49] Kouri T, Kairisto V, Virtanen A, Uusipaikka E, Rajamäki A, Finneman H, et al. Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. *Clin Chem*. 1994; 40:2209-2215.

[50] Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem*. 1998;44:622-631.

[51] Laboratory Accreditation Program Automated/General Chemistry Checklist. AGC.20200. Northfield, IL: College of American Pathologists; 2001.

Additional References

Deming WE. *Out of the Crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study; 1986.

England JM, Rowan RM, van Assendelft OW, et al. Guidelines for organisation and management of external quality assessment using proficiency testing. *Int J Hematol.* 1998;68:45-52.

Appendix A. Procedure to Determine Allowed Differences Between Laboratories X and Y in a Split-Sample Program

The following discussion assumes that the laboratory has sufficient experience with the method to know the capabilities of the method, including internal precision (i.e., repeatability, variance) and biases, if any, across the entire clinical range of test results. If these are not known, then there is no suitable way to validate method performance by split-sample procedures.

In external split-sample procedures, usually the outside laboratory(ies) will use the same method, but this is not necessary as long as the relationship between the methods is known. The laboratories should understand any differences (including differences in specificity) between the methods. Refer to the most current edition of NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples* for guidance in determining the relative difference between two methods. The most current editions of NCCLS documents EP7—*Interference Testing in Clinical Chemistry* and EP21—*Total Error for Clinical Laboratory Methods* provide information on estimating differences in specificity.

It is important that the laboratories involved have agreed-on criteria for test assessment before initiating the split testing. This agreement should include the following:

- test methods and numbers of tests to use;
- criteria for determining agreement;
- whether assessment will be at specific levels or across a range of levels; and
- procedures to resolve disagreements, which may include:
  - reruns by one or both laboratories;
  - whether one of the laboratories is considered the reference for the other(s); and
  - whether to consult a third laboratory.

The laboratories should retain the results so that eventually comparisons will cover a wide range of concentrations. The laboratories can assess agreement across this range by plotting results on a two-dimensional graph with the result from the laboratory of highest authority (if either[k]) on the horizontal (X) axis and the other laboratory on the vertical (Y) axis, and a line of perfect agreement (Y=X) drawn in the body of the graph. The laboratories may discover that their results tend to disagree in a predictable way (that perhaps is correctable), or that disagreements are random, but are scattered on both sides of the line of perfect agreement. Visual assessment should be adequate to identify predictable trends, or least squares regression analysis may be used to assist in the determination.[l]

It is important that the capabilities of the assessment method are consistent with the uncertainty that is desired. For example, if at a certain level the internal QC data show typical repeatability (coefficient of variation) of +/- 2%, then the laboratory can be sure only that real changes of 4% or more will be reflected in test results on single samples. If for clinical reasons it is important to detect a change of 3%,

[k] Traditionally the horizontal axis is used for the independent variable and the vertical axis for the dependent variable. If one of the laboratories is considered of "higher authority," then in this context it is considered the comparison laboratory, and its values should be placed on the horizontal axis.

[l] Technically, orthogonal regression should be used, since both sets of results have error in them, but this is not necessary for the purpose at hand, and orthogonal regression procedures are not widely available. Least squares regression can be used for basic descriptive purposes if the test samples cover a broad range of concentrations. Differences between least squares regression and orthogonal regression are minimal if the tests cover a broad range of concentrations.

then the method is not adequate (it is said to have insufficient power to detect changes of 3%). In such a case, the laboratory can run the test in duplicate or triplicate and use the mean as the test result. Because the internal QC CV of a single assay in this example is 2%, the mean of two test results has a CV of 1.4%, and thus the use of duplicate testing provides the laboratory with reasonable power to detect real changes of 3% or more.

The following example demonstrates the use of a formula to calculate the confidence interval for the difference between results from two procedures or two laboratories, based on data that have been accumulated over several years of split-sample testing. For this analysis to be valid, the testing systems in each laboratory must have been stable over the period of data accumulation.

Two laboratories run a test for serum antibody IgZ1. Occasionally they send samples to each other for validation of the performance of the test, and each laboratory runs each test in duplicate. Table A1 shows the results for 18 samples tested by the two laboratories. For convenience, they are ordered by concentration.

Table A1. IgZ1 Results from Laboratories X and Y. Results on duplicate samples for 18 patients, arranged in ascending antibody levels. Differences between laboratory averages and the allowed difference are noted.

Sample	Laboratory X		Laboratory Y		Averages from X and Y			
	Rep1	Rep2	Rep1	Rep2	Average X	Average Y	Diff. X-Y	Allowed[m]
1	790	800	630	577	795.0	603.5	191.5	290.3
2	861	905	543	664	883.0	603.5	279.5	322.4
3	1051	1174	725	784	1112.5	754.5	358.0	406.2
4	1846	1846	1419	1632	1846.0	1525.5	320.5	674.0
5	1894	1820	1974	2363	1857.0	2168.5	-311.5	678.0
6	2014	2270	1550	1451	2142.0	1500.5	641.5	782.1
7	2484	2460	1640	1416	2472.0	1528.0	944.0	902.5
8	2405	2684	2096	2535	2544.5	2315.5	229.0	929.0
9	2560	3065	2181	2340	2812.5	2260.5	552.0	1026.9
10	2612	3065	1961	1887	2838.5	1924.0	914.5	1036.4
11	5755	5585	8415	8166	5670.0	8290.5	-2620.5	2070.2
12	5812	5812	6424	7171	5812.0	6797.5	-985.5	2122.0
13	8705	8473	6619	5989	8589.0	6304.0	2285	3135.9
14	9116	8671	10591	9875	8893.5	10233.0	-1339.5	3247.1
15	10029	9880	10697	10486	9954.5	10591.5	-63.0	3634.5
16	11736	12585	9393	10591	12160.5	9992.0	2168.5	4439.9
17	12554	12807	13874	13509	12680.5	13691.5	-1011.0	4629.7
18	14473	14705	15106	15174	14589.0	15140.0	-551.0	5326.6

Assume that the laboratories have assessed their tests in-house by running routine replicates on patient samples. Laboratory X is more experienced with the test, and is acknowledged as the more reliable of the two laboratories for this test. The standard deviations from the in-house repeat tests show that in Laboratory X the repeatability (precision) CV is approximately 10% across the range of concentrations tested. In Laboratory Y the repeatability is approximately CV 12%. Further, a published study suggests that interlaboratory variability is approximately CV 15%, which will be used as the clinically acceptable agreement.

[m] Calculated as shown by Formula (1) below.

Therefore we assume the following:

A = antibody level measured by Laboratory X
σ_X^2 = repeatability variance for Laboratory X ≈ $(.10*A)^2$
σ_Y^2 = repeatability variance for Laboratory Y ≈ $(.12*A)^2$
σ_I^2 = interlaboratory variance ≈ $(.15*A)^2$
n_X = n_Y = number of replicates from each laboratory (can be 1)
α = confidence level
$z_{1-\alpha/2}$ = percentile of the normal distribution corresponding to the level $1-\alpha/2$

Note that repeatability variances can come from QC data or published method capability sources as in this example.[n] If QC data are used, σ_X^2 and σ_Y^2 equal the square of the standard deviation of the applicable internal QC data. Repeatability variances can also be estimated from the split-sample data, if multiple replicates are used. The laboratory repeatability is calculated as the pooled proportional difference between replicates [(Rep1-Rep2)/((Rep1+Rep2)/2))]. This ratio is squared, the squared terms are then summed over all samples, and the sum is divided by the number of samples minus 1 to give the σ_X^2 or σ_Y^2, respectively. Similarly, we could estimate the interlaboratory variance by first calculating the sum of squared differences, divided by the number of samples minus 1. This is the variance of the differences; the interlaboratory variance would then be estimated by subtracting the combined repeatability variance ($S^2r = (s_x^2 + s_y^2/2)$). If the repeatability variance exceeds the variance of differences, the interlaboratory variance is set at zero. If interlaboratory variance is unknown at the initiation of a split-sample comparison program, or if it can be assumed that the laboratories should produce equivalent results, interlaboratory variance can be set at 0.

If the estimates of variability had not been available elsewhere, they could have been estimated from the data in Table A1.

Allowed limits are determined with formula XYZ, as follows:

[Formula (1)]

Allowed difference $D = z_{1-\alpha/2}\sqrt{\left(\sigma_1^2 + \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)}$

If the difference between any pair of individual results lies within the confidence interval, the difference is not statistically significant, and the results can be considered equivalent for the individual sample.

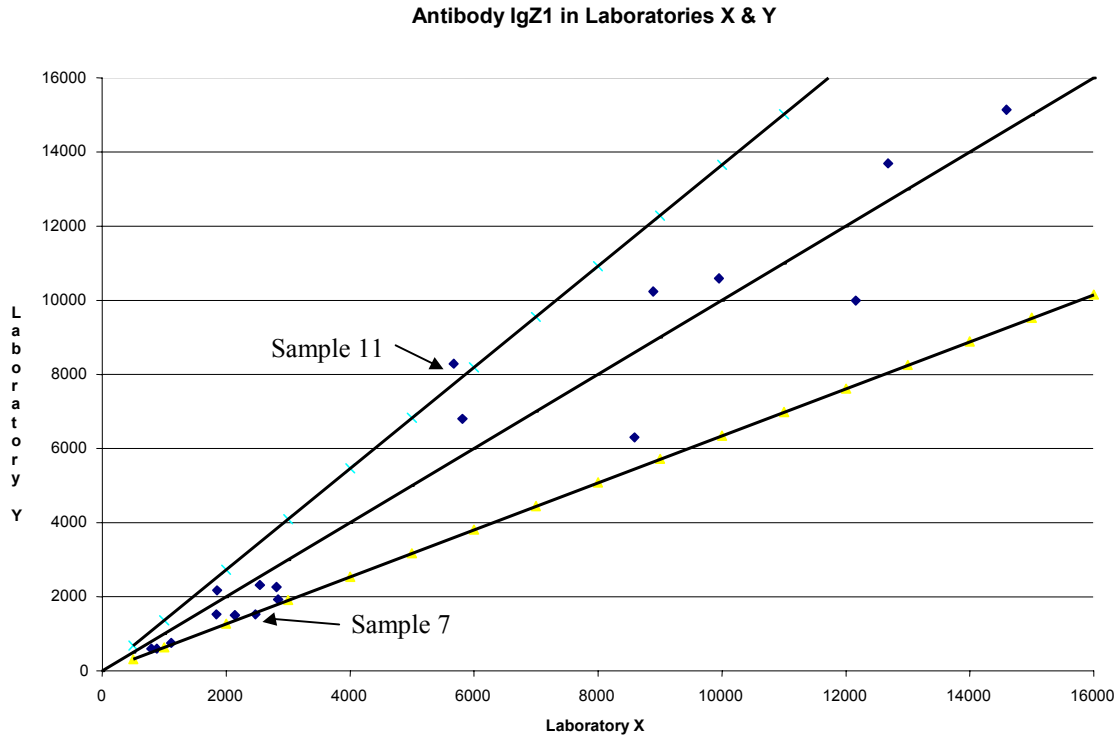
At, for example A = 2470:

σ_X^2 = $(.10*2470)^2$ = 61009
σ_Y^2 = $(.12*2470)^2$ = 87853
σ_I^2 = $(.15*2470)^2$ = 137270
n_X = n_Y = 2
α = confidence level = .95
$z_{1-\alpha/2}$ = 1.96
D = 1.96 (square root of) [137270+61009/2+87853/2]
D = 1.96 (460) = 902

[n] Refer to the first part of Section 7 for further suggestions of ways to define limits of acceptability for variance in AAPs.

So at this level of antibody, a difference of up to 900 would be expected. Note that Sample 7 was at this level of antibody, and the observed difference exceeds the allowed limit. Sample 11 shows an even greater difference.

The following chart shows the averages plotted, with the line of equality and the allowed differences. One can see that Sample 7 barely exceeds the limits, while Sample 11 is quite far from the expected difference. Both samples should be investigated for anomalies.



The middle line represents perfect agreement; the upper and lower lines represent the acceptable difference for results.

Calculations are not shown here, but for these samples we would get estimates of variance as follows:

s_x = 0.08 (compared with .10 from internal QC data)
s_y = 0.11 (compared with .12 from internal QC data)
s_l = 0.21 (compared with .15 assumed from literature report)

Appendix B. Method for Determination of Agreement with a Reference Method

If the method of the outside laboratory can be considered to be a reference method, the following calculation of confidence intervals to determine agreement can be used:

The formula for a confidence interval (CI) at any level L can be expressed as:

[Formula (2)]

$$CI = L + B \pm Z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}$$

where:

$z_{1-\alpha/2}$ = percentile of the standard normal distribution for stated confidence level 1-α (e.g. 95% or 99%)
L = concentration level of interest (usually the level tested in the external verification experiment)
B = known bias or difference between test method and reference method (often B=0)
σ^2 = variance of the test method at level L
n = number of replicates used in verification procedure

If the external test result lies within the confidence interval, then the method is considered to have passed the verification at that level.

Appendix C. Statistical Evaluation of Qualitative Split-Sample Results*

Tests that have binary results—e.g., positive or negative—pose a particular problem in split-sample testing, because of the likelihood of chance agreement between two laboratories. To evaluate the contribution of chance to agreement between two sets of data, the kappa statistic may be used. The kappa statistic compares observed agreement with agreement that might be expected by chance. Kappa values range from +1 (complete agreement) through 0 (no agreement beyond that expected by chance) to –1 (complete disagreement—which in the laboratory setting suggests systematic reversal of results, perhaps by clerical or programming error). Kappa values above 0.8 can be considered excellent analytic agreement, and those between 0.6 and 0.8 can be considered to be reasonable agreement. With samples of 20 or more, kappa larger than 0.5 is statistically significant, indicating that agreement is not entirely due to chance.

The kappa statistic is calculated as follows for two laboratories, A and B:

Kappa=(observed agreement – chance agreement)/(1-chance agreement)

Chance agreement=(proportion of negative results, Laboratory A) x (proportion of negative results, Laboratory B) + (proportion of positive results, Laboratory A) x (proportion of positive results, Laboratory B)

For example, Laboratories A and B have performed split-sample testing on 29 specimens over the past several years, with the following result:

	Laboratory A		**Total**
	Negative	Positive	
Laboratory B			
Negative	9 (31.0%)	5 (17.2%)	14 (48.3%)
Positive	1 (3.5%)	14 (48.3%)	15 (51.7%)
Total	10 (34.5%)	19 (65.5%)	29 (100%)

Using the data from the table (and multiplying the percentage figures by .01):

observed agreement = (9+14)/29 = .793.

chance agreement = (.483 x .345) + (.517 x .655) = .505.

Thus, kappa = (.793 - .505)/(1 - .505) = .58.

With this size sample, kappa = .58 indicates that the agreement between laboratories is not extremely strong, but is greater than could have occurred by chance alone.

Kappa is useful for comparing agreement on different tests, agreement between different laboratories, or tracking changes over time. This can assist in finding causes for the disagreement.

*Adapted from Dunn G, Everitt BS. *Clinical Biostatistics: An Introduction to Evidence Based Medicine.* London: Edward Arnold; 1995. Reproduced by permission of Arnold Publishers.

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

GP29-P: *Validation of Laboratory Tests When Proficiency Testing is Not Available; Proposed Guideline*

Section 5

1. I would suggest that it is made clear in Section 5 that these are categories of tests with examples for which PT may not be available, but that the list is not all-inclusive. We have found no PT available for vitamin A or carotene, among others and I'm sure there are many drugs for which PT is also unavailable.

- **The list was not intended to be all-inclusive, but a clarifying statement has been added. The suggested analytes have been added to the list.**

Section 6

2. 2nd paragraph, final sentence. It would be appropriate to mention that laboratories should consider national guidance as well as the institution's requirements with respect to informed patient consent and maintenance of patient confidentiality.

- **The subcommittee believes that the wording "their institution's requirements for informed patient consent…etc." is adequate, because institutions should take into account local, regional, and national guidelines as well as laws and regulations. Additional language regarding national guidelines would be redundant.**

Section 6.5

3. Penultimate paragraph. It would be useful to stress the importance of individual laboratories establishing their own reference ranges for the population they serve, rather than relying on reference ranges quoted by instrument manufacturers.

- **Language addressing this issue has been added.**

Section 7.1

4. Critiques of the sections on statistical analysis of split-sample and patient data were specifically requested. I have spent some time studying these and find them logical and straightforward. The example in paragraph 7.1 will be particularly helpful.

- **The subcommittee appreciates the input (and the example was moved to Appendixes A and B).**

Section 7.1.1

5. The statistical treatment of split-sample testing should be removed to an appendix. The example given should be more clearly labeled as an example rather than a prescription for the number of samples to be split. The appendix should also offer an approach to the analysis of qualitative analysis split between two laboratories.

- **The example has been moved to Appendixes A and B. Text has been added to emphasize that it is the responsibility of the laboratory to determine the number of samples to be tested. An approach to analysis of data from qualitative split-sample testing has been added in Appendix C.**

6. The target value in split-sample testing may be based upon clinical criteria. For example, suppose that a qualitative test had three possible results: reactive, indeterminate, and nonreactive. The lab knows that clinicians will treat only on the basis of a reactive result. Therefore, split-sampling that results in an indeterminate result from one lab and a nonreactive from another may be considered a verification of accuracy.

- **The document in fact states that clinical criteria may define limits of acceptability of results of alternative assessment procedures (Section 6, second paragraph).**

7. The document should address the number and periodicity of analytic verifications without being prescriptive. A statement like, "Verification of two analyses every six months for esoteric tests is adequate," is suggested.

- **This point is covered in the document. See footnote f: "In the U.S., CLIA '88 requires that AAPs be performed twice yearly.[12] This is adequate for most situations." The language has been edited slightly.**

Summary of Delegate Comments and Committee Responses

GP29-A: *Validation of Laboratory Tests When Proficiency Testing is Not Available; Approved Guideline*

General

1. Overall, this is a very good document for medical technologists. However, this document would not be used by staff in a physician's office laboratory (POL), because it would be too technical and assumes a knowledge base most staff in a POL do not have.

- **The document was crafted to be useful to a wide constituency of laboratorians, and so technical material was included. The subcommittee believes that some of the concepts presented in the document—split-sample testing, re-evaluation of morphologic analysis, direct observation by supervisors—are easily understood in their essence by nontechnical personnel, and may be useful to them. However, the document was not intended to serve as a primer in quality management of laboratory testing.**

Title

2. The title of this document misuses an internationally defined term, i.e., validation. GP29 suggests that proficiency testing (PT) is used to validate a test method. PT does not validate a test method. PT is used to monitor performance after it has been validated. I believe the title should probably read, "Assessment of Laboratory Tests When Proficiency Testing is Not Available."

   ISO has clear definitions of validation and verification, which have been accepted by FDA, the IVD industry, ILAC and others who use NCCLS documents. Furthermore, test method validation has very specific meaning and expectations in FDA-regulated industries, including the IVD industry, and there are guidelines on validation of analytical methods that have been published by the International Conference on Harmonization that have been widely adapted for use in the IVD industry (ICH Q2A and Q2B - available on the FDA website).

   In a nutshell, test method validation means demonstrating with objective evidence that the method is suitable for its intended use. The process of validation is not trivial. It requires clear definition of the requirements of the test method (in the case of the clinical lab, they would depend on the medical use of the results), and demonstration that the analytical performance characteristics (bias, imprecision, specificity, linearity, etc.) satisfied the medical need. Validation is done after the performance of the method as been characterized and before the method is used to report results. Typically, validation studies require an approved protocol with a statistically valid study design and predetermined acceptance criteria, to ensure the validation is objective and scientifically sound.

   Once a method has been validated, it is monitored by internal quality control and external PT to verify its performance remains stable. This can be viewed as a continuation of the validation, but it is not the validation.

   The terms validation and verification have been loosely used in the clinical lab, partly due to their introduction by CLIA without a clear understanding of their meaning. NCCLS can help rectify that, but if an NCCLS document is published that implies a method can be validated by running PT samples, or by the alternative strategies outlined in GP29, NCCLS will just add to the confusion in the clinical labs and lose credibility among the industry sector and the international community.

- **Based on input from the NCCLS Harmonization Group, the subcommittee has replaced the term "validation" with "assessment."**

3. The document would be better if it described a more coherent process for (1) transferring a validated method from one lab (or from a manufacturer) to another lab, and (2) monitoring its performance via PT or an alternative performance assessment procedure. NCCLS might want to consider a guideline for clinical labs that describes how to conduct a transfer study when bringing a validated method into the lab. That could help eliminate a lot of confusion and unnecessary work, since commercial systems sold by manufacturers have already been validated and do not require revalidation by the lab, only a demonstration that they are performing as intended.

- **Regarding (1), the scope of the document does not include transfer studies or initial validation when a method is introduced to a laboratory. These issues are addressed in NCCLS documents EP9—*Method Comparison and Bias Estimation Using Patient Samples;* EP12—*User Protocol for Evaluation of Qualitative Test Performance;* and EP15—*User Demonstration of Performance for Precision and Accuracy*. Regarding (2), the scope of the document does include monitoring test performance by alternative methods when PT is unavailable, but not monitoring performance by PT *per se*.**

Foreword

4. The methods described in this document are not *sufficient* to validate tests. They are options to include in the rigorous process needed for validation. (See Comment 6.) I recommend editing the sentence as follows: "…methods to help validate tests…" There is also no distinction between validation of analytical performance (achieving "trueness") from clinical performance, which is meeting the intended use of the method. These might not be the same.

- **The subcommittee did not intend to imply that PT or alternative procedures are sufficient *in themselves* to validate test performance in the sense of "demonstrating with objective evidence that the method is suitable for clinical use." (ISO definition)**

  The wording of the Foreword has been modified for clarity.

  The term "validation" has been replaced by "assessment" in the document (see response to Comment 2 above).

  The document does not address assessment of "trueness" or accuracy except where specifically noted (for example, Section 5, fifth bullet). The document primarily focuses on subsets of clinical performance assessment which are addressed by PT or alternative procedures, such as error detection and interlaboratory comparability. Refer to Section 5 of the document.

Introduction

5. The fourth sentence of the first paragraph gives the impression that PT is a validation of truth and that PT is a necessary requirement for validation. Neither, of course, is true. I recommend editing the sentence as follows: "…development of alternative procedures that can be used to help validate tests if PT is not available or does not fulfill the need of the validation plan."

- **As noted in the response to Comment 2 above, the term "validation" has been replaced by "assessment." The sentence has been edited to read, "This document offers guidance to clinical laboratories in the development of alternative assessment procedures (AAPs) when PT is not available."**

Section 4.1 (formerly Section 3.1)

6. Quality systems (Quality System regulation and ISO 9000) have a definition for "validation" that should be used. It is more descriptive and provides consistency with accepted standards (i.e., Validation: Confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use can be consistently fulfilled).

- **The term "validation" has been replaced by "assessment," as described in the response to Comment 2 above.**

Section 5 (formerly Section 4)

7. Some PT programs use native specimens or specimen pools. Therefore, we recommend rewording the sixth paragraph to read, "AVPs may use patient specimens which have certain advantages over manufactured materials frequently used in PT programs."

- **The suggested change has been made.**

8. I believe the following sentence is misleading: "Unlike PT, a AVPs may use patient specimens, which have certain advantages over the manufactured materials used in PT." Our EQA program uses some patient specimens (for smaller numbers of users). Suggest revising to read, "Unlike THE MAJORITY OF PT, AVPs may use patient specimens, which have certain advantages over the manufactured materials USUALLY used in PT."

- **See response to Comment 7 above.**

Section 7 (formerly Section 6)

9. Although using internal QC data for establishing goals is valid, there are two issues with including this recommendation without further qualification. (1) It is unlikely that a new test would have sufficient data to use to establish robust goals that would include sources of variability that are known only after longer use, e.g., calibration, lot, operator, stability, etc. (2) QC data provide a limited picture of test capability (see Comment 17).

   I recommend that the subcommittee revisit this paragraph. Defining limits for acceptance *a priori* is a requirement for quality systems (and clearly indicated). However, it is dangerous to establish limits (specifications) for a parameter, assumed to be "total allowable error," which is assumed to be the objective of PT when internal QC for a new test would not "capture" other components of reproducibility. Also, it is not at all capable of providing objective evidence of test specificity – freedom from interference. Consult WH Lawton, EA Sylvestre, BJ Young-Ferraro. Statistical comparison of multiple analytic procedures: application to clinical chemistry; *Technometrics*. 1979; 21:397-406.

- **Regarding the first point, a caveat has been added to Section 7 as follows: "provided that sufficient QC data exist." Regarding the second point, the subcommittee agrees that QC data address only one aspect of a method's capability. QC performance is used in this section as an example of one of several data sources that can be used to establish acceptance criteria for AAPs. The change of terminology in the document from AVP to AAP should clarify that these techniques are not intended to comprehensively validate test performance.**

Section 7.1.1 (formerly Section 6.1.1)

10. What does the committee suggest on how to provide a split (purulent) sputum sample to another laboratory for validation?

- **The response to the comment is dependent on the test of being studied. If the test of interest is culture, for example, a single sample must be split and evaluated immediately, in real time, by both laboratories. If the test under consideration is a cytologic evaluation, an aliquot of preserved sample can be taken and evaluated by the second laboratory when convenient to do so; alternatively, the first laboratory can simply send slides to the second laboratory for review after the first laboratory has completed its evaluation. This last alternative does not include the evaluation of slide preparation and staining in the split-sample study, however.**

11. While I do not disagree with the statement that "…split-sample testing (procedures) do not evaluate trueness (bias) per se unless the outside laboratory uses a reference method", this same argument could be used against most proficiency testing materials, which derive their acceptable ranges by a consensus of results from several labs. In fact, for many clinical chemistry tests, the manufacturers calibrate their instruments to a reference method or reference material. Because of that, one could argue that between-laboratory comparisons of measurements on split-sample do evaluate accuracy. We have used weekly whole blood split-sample comparisons for many years for our blood gas/electrolyte/glucose analyzers. My experience has been that it is one of the most reliable indicators of analytical deviations.

- **As the subcommittee reads the above comment, there is no disagreement with the document. To clarify the issue, the statement in Section 7.1.1 is changed to read, "split-sample procedures…do not evaluate trueness …unless the outside laboratory uses a method that is calibrated to a reference method or reference material."**

Section 7.3 (formerly Section 6.3)

12. Commutable is too stringent a requirement. There is a new ISO/EN proposed standard on calibrator traceability that could be referenced to satisfy this need as well. I recommend editing the sentence as follows: "documented to be commutable with patients' specimens or are traceable to a reference material or procedure to confirm correct performance…" (Reference prEN ISO 17511).

- **The subcommittee agrees with the comment and has made the recommended change.**

Section 7.5.3 (formerly Section 6.5.3)

13. Delete this section. Delta checking is not an effective mechanism for method validation. It is most effective for identifying sample mix-ups, but is not sufficiently sensitive to identify other issues. If used, it has the significant risk of providing a false sense of security that the method is meeting its intended use (the goal of test validation).

- **The section is retained, because circumstances might exist in which delta checking is the only feasible alternative assessment method. The subcommittee agrees that delta checking is not a robust means of test assessment. See the responses to Comments 2 and 4 above, which clarify that the document does not address comprehensive validation of test methods.**

Sections 7.6, 7.9, and 7.10 (formerly Sections 6.6, 6.9, and 6.10, respectively)

14. While the rest of the document describes procedures that are generally applicable to nearly all laboratory tests, these sections apply to very specific tests. I would guess that these topics have been

well debated before, so my only suggestion would be to group these all under one section with the title to describe their application to specific limited-scope tests.

- **The subcommittee agrees that these topics have been discussed before, but feels that the document is easy to follow as it is presently organized.**

Section 7.8 (formerly Section 6.8)

15. The sense of this section is false as written. I agree this has limited value, but when needed, it is invaluable. Many new tests can only be assessed as effective against its ability to support diagnosis or reasonably reflect the affects of therapy. Rewrite to provide guidance when clinical correlation studies are useful, while pointing out the limitations. Eliminate inference that it cannot be used for validation "because of the imperfect correlation of clinical events to laboratory results…" Few tests (that are used for specific diagnosis or disease monitoring) are "perfect;" that is the value of clinical specificity and sensitivity studies.

- **The point of this section is that clinical correlation studies are useful for certain tests—generally, tests that are evaluated against a threshold value for the diagnosis of a specific disorder, the presence of which can be verified independently with a reasonable amount of effort. Examples include serum CK-MB for the diagnosis of acute myocardial infarction and serum amylase for the diagnosis of acute pancreatitis. Clinical correlation is not as useful for a test which may be abnormal in many disorders, or when the presence or absence of the disorder(s) under study is not amenable to timely independent verification. The wording in Section 7.8 has been changed to clarify this point.**

Section 8 (formerly Section 7)

16. There is only one substantive recommendation, which is if a definitive diagnosis is available. Yet, even with the caveats mentioned, there are many tests that have used clinical sensitivity and specificity for validation, even when there is no " gold standard" for definitive diagnoses. (See Comment 14. Appendix C is very helpful and can be used to mollify the concerns of Section 7.)

- **As the subcommittee reads the comment, the reviewed does not appear to disagree with the discussion in Section 8.**

Appendix A (formerly Appendix A, Part 1)

17. The advice in the fourth paragraph is true, but using QC results does not satisfy the need of validating a method to assure that the total allowable error can be achieved. See Comment 9. An alternative is to budget allowable error into bias, precision (repeatability and reproducibility), and method specificity (see article referenced in Comment 9). This approach, although easier to verify performance, is much more difficult to develop.

- **The techniques presented in this document are intended to provide ongoing assessment of the validity of laboratory tests. The techniques are not intended to "satisfy the need of validating a method to assure that the total allowable error can be achieved." (See responses to Comments 2 and 4 above). Aspects of method validation are addressed in the most current editions of NCCLS documents EP5—*Evaluation of Precision Performance of Clinical Chemistry Devices*; EP6—*Evaluation of the Linearity of Quantitative Analytical Methods*; EP7—*Interference Testing in Clinical Chemistry*; and EP9—*Method Comparison and Bias Estimation Using Patient Samples*. Considerations in the estimation of total error are contained in the current edition of NCCLS document EP21—*Total Error for Clinical Laboratory Methods*.**

18. This is a standard method for doing this analysis, but it does not account for one of the more common sources of difference between two methods: specificity [often demonstrated in a method's sensitivity to matrix effects]. Add a comment that this is useful, but the estimate is based on repeatability and reproducibility only. Note Comment 9 about the probable limited data available for a new method.

- **The procedure in Appendixes A and B states that the laboratories either use the same method, or that the methods must have a "known relationship." This statement implies that there are no specificity differences. A comment on this subject has been added to the document.**

Appendix B (formerly Appendix A, Part 2)

19. See concern mentioned in Comment 17 above. Method validation must include an understanding of specificity.

- **See the response to Comment 18 above. Also, as noted above (see responses to Comments 2 and 4), the document does not address initial test validation, but rather ongoing assessment.**

Summary of Comments and Subcommittee Responses

20. The response to Comment 7 is not sufficiently broad. I view this document as helpful for several purposes, including for use by laboratories that develop their own tests and need to perform routine verification of performance. Include a statement in footnote f with regard to other reasons to verify performance. (Note: verification is defined as "confirmation by examination and provision of objective evidence that specified requirements have been fulfilled." This is an acceptable means of confirming that the method still meets its intended use.) Reasons to either re-validate (probably not required) or to verify conformance to specifications can include such things as raw material changes, e.g., a new antibody source.

- **The document discusses assessment of test performance when PT is not available. Verification may be needed for other reasons, but these lie outside the scope of the document. See the responses to Comments 2 and 4 above.**

Related NCCLS Publications*

EP9-A2 **Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition (2002).** This document addresses procedures for determining the bias between two clinical methods or devices, and for the design of a method comparison experiment using split patient samples and data analysis.

EP14-A **Evaluation of Matrix Effects; Approved Guideline (2001).** This document provides guidance for evaluating the error or bias in analyte measurements that is due to the sample matrix (physiological or artificial) when two analytical methods are compared.

GP27-A **Using Proficiency Testing (PT) to Improve the Clinical Laboratory; Approved Guideline (1999).** This guideline provides assistance to laboratories in using proficiency testing as a quality improvement tool.

HS1-A **A Quality System Model for Health Care; Approved Guideline (2002).** This document provides a model for healthcare service providers that will assist with implementation and maintenance of effective quality systems.

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).

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

NOTES

NOTES

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 1-56238-479-1