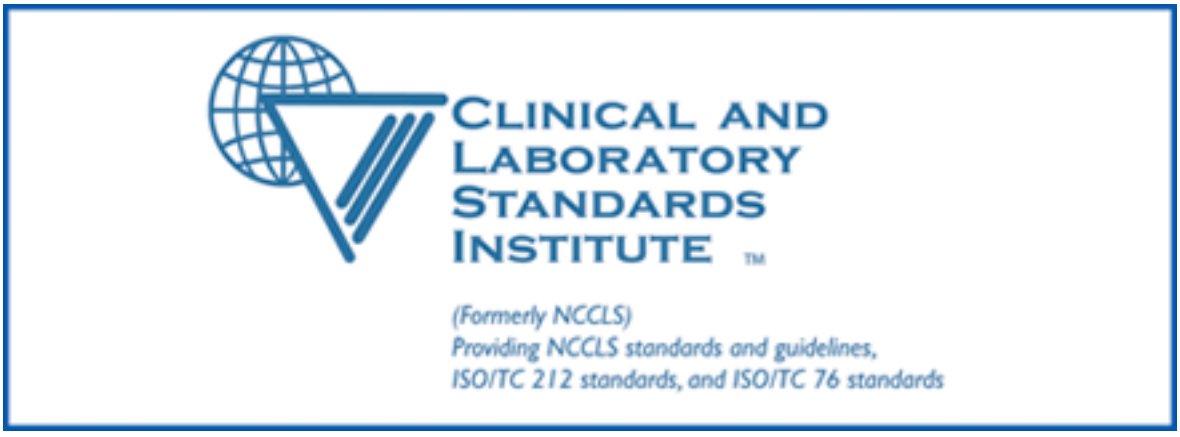
EP19-R
Vol. 22 No. 10

Replaces EP19-P
Vol. 20 No. 16

A Framework for NCCLS Evaluation Protocols; A Report



This report describes the different types of performance studies that are conducted to evaluate clinical assays.

A report developed by the NCCLS Area Committee on Evaluation Protocols

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.

A Framework for NCCLS Evaluation Protocols; A Report

Abstract

This report describes the different types of performance studies that are conducted to evaluate clinical assays: namely, validation, verification, and demonstration studies. Each evaluation protocol guideline is categorized with respect to its evaluation type. The tasks that comprise the path of workflow, which takes place when an assay is considered for use by a laboratory, are listed and explained. A recommendation for the data analysis portion of the evaluation is discussed, whereby the use of confidence intervals over hypothesis testing is recommended.

NCCLS. *A Framework for NCCLS Evaluation Protocols; A Report*. NCCLS document EP19-R (ISBN 1-56238-464-3). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.

EP19-R
ISBN 1-56238-464-3
ISSN 0273-3099

A Framework for NCCLS Evaluation Protocols; A Report

Volume 22 Number 10

Carl C. Garber, Ph.D.
Jan S. Krouwer, Ph.D.
R. Neill Carey, Ph.D.
Larry W. Clark
Patricia E. Garrett, Ph.D.
John W. Kennedy
Martin Harris Kroll, Ph.D.
Fred D. Lasky, Ph.D.
Henry T. Lee, Jr.
Jacob B. Levine, M.B.A.
Herbert K. Naito, Ph.D., M.B.A.
Richard B. Passey, Ph.D.
David L. Phillips

Suggested Citation

(NCCLS. *A Framework for NCCLS Evaluation Protocols; A Report.* NCCLS document EP19-R [ISBN 1-56238-464-3]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.)

Proposed Guideline
July 2000

Report
Approved by the Board of Directors
May 2002

Published
June 2002

ISBN 1-56238-464-3
ISSN 0273-3099

Committee Membership

Area Committee on Evaluation Protocols

Jan S. Krouwer, Ph.D. **Chairholder**	**Krouwer Consulting** **Sherborn, Massachusetts**
Daniel W. Tholen, M.S. **Vice-Chairholder**	**Statistical Services** **Traverse City, Michigan**
Carl C. Garber, Ph.D.	Quest Diagnostics Assurance Teterboro, New Jersey
Henk M.J. Goldschmidt, Ph.D.	Tilburg, The Netherlands
Martin Harris Kroll, M.D.	Dallas Veterans Affairs Medical Center Dallas, Texas
Kristian Linnet, M.D., Ph.D.	Psychiatric University Hospital Risskov, Denmark
Kristen Meier, Ph.D.	FDA Center for Devices/Rad. Health Rockville, Maryland
Max Robinowitz, M.D.	FDA Center for Devices/Rad. Health Rockville, Maryland

Advisors

R. Neill Carey, Ph.D.	Peninsula Regional Medical Center Salisbury, Maryland
Patricia E. Garrett, Ph.D.	BBI Clinical Laboratories New Britain, Connecticut
John W. Kennedy	Medstat Consultants Palo Alto, California
Jacob (Jack) B. Levine, M.B.A.	Bayer Corporation Tarrytown, New York
Jennifer K. McGeary, M.T.(ASCP), M.S.H.A. *Staff Liaison*	NCCLS Wayne, Pennsylvania
Patrice E. Polgar *Editor/Project Manager*	NCCLS Wayne, Pennsylvania
Donna M. Wilhelm *Assistant Editor*	NCCLS Wayne, Pennsylvania

Active Membership
(as of 1 April 2002)

Sustaining Members

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

Professional Members

AISAR-Associazione Italiana per lo Studio degli
American Academy of Family Physicians
American Association for Clinical Chemistry
American Association for Respiratory Care
American Chemical Society
American Medical Technologists
American Public Health Association
American Society for Clinical Laboratory Science
American Society of Hematology
American Society for Microbiology
American Type Culture Collection, Inc.
Asociación Española Primera de Socorros (Uruguay)
Asociacion Mexicana de Bioquimica Clinica A.C.
Assn. of Public Health Laboratories
Assoc. Micro. Clinici Italiani-A.M.C.L.I.
British Society for Antimicrobial Chemotherapy
CADIME-Camara De Instituciones De Diagnostico Medico
Canadian Society for Medical Laboratory Science—Société Canadienne de Science de Laboratoire Médical
Clinical Laboratory Management Association
COLA
College of American Pathologists
College of Medical Laboratory Technologists of Ontario
College of Physicians and Surgeons of Saskatchewan
ESCMID
Fundación Bioquímica Argentina
International Association of Medical Laboratory Technologists
International Council for Standardization in Haematology
International Federation of Clinical Chemistry
Italian Society of Clinical Biochemistry and Clinical Molecular Biology
Japan Society of Clinical Chemistry
Japanese Committee for Clinical Laboratory Standards
Joint Commission on Accreditation of Healthcare Organizations
National Academy of Clinical Biochemistry
National Society for Histotechnology, Inc.
Ontario Medical Association Quality Management Program-Laboratory Service
RCPA Quality Assurance Programs PTY Limited
Sociedade Brasileira de Analises Clinicas
Sociedade Brasileira de Patologia Clinica
Sociedad Espanola de Bioquimica Clinica y Patologia Molecular
Turkish Society of Microbiology

Government Members

Association of Public Health Laboratories
Armed Forces Institute of Pathology
BC Centre for Disease Control
Centers for Disease Control and Prevention
Centers for Medicare & Medicaid Services/CLIA Program
Centers for Medicare & Medicaid Services
Chinese Committee for Clinical Laboratory Standards
Commonwealth of Pennsylvania Bureau of Laboratories
Department of Veterans Affairs
Deutsches Institut für Normung (DIN)
FDA Center for Devices and Radiological Health
FDA Center for Veterinary Medicine
FDA Division of Anti-Infective Drug Products
Health Care Financing Administration/CLIA Program
Health Care Financing Administration
Iowa State Hygienic Laboratory
Massachusetts Department of Public Health Laboratories
National Association of Testing Authorities – Australia
National Center of Infectious and Parasitic Diseases (Bulgaria)
National Health Laboratory Service (South Africa)
National Institute of Standards and Technology
New York State Department of Health
Ohio Department of Health
Ontario Ministry of Health
Pennsylvania Dept. of Health
Saskatchewan Health-Provincial Laboratory
Scientific Institute of Public Health; Belgium Ministry of Social Affairs, Public Health and the Environment
Swedish Institute for Infectious Disease Control
Thailand Department of Medical Sciences

Industry Members

AB Biodisk
Abbott Laboratories
Abbott Laboratories, MediSense Products
Acrometrix Corporation
Advanced Medicine, Inc.
Agilent Technologies, Inc.
Ammirati Regulatory Consulting
Anaerobe Systems
Asséssor
AstraZeneca
Aventis
Axis-Shield POC AS
Bayer Corporation – Elkhart, IN

Bayer Corporation – Tarrytown, NY
Bayer Corporation – West Haven, CT
Bayer Medical Ltd.
BD
BD Biosciences – San Jose, CA
BD Consumer Products
BD Diagnostic Systems
BD Italia S.P.A.
BD VACUTAINER Systems
Beckman Coulter, Inc.
Beckman Coulter, Inc. Primary Care Diagnostics
Beckman Coulter K.K. (Japan)
Bio-Development SRL
Bio-Inova Life Sciences International
Bio-Inova Life Sciences North America
BioMedia Laboratories Sdn Bhd
BioMérieux (NC)
bioMérieux, Inc. (MO)
Biometrology Consultants
Bio-Rad Laboratories, Inc.
Bio-Rad Laboratories, Inc. - France
Biotest AG
Blaine Healthcare Associates, Inc.
Bristol-Myers Squibb Company
Canadian External Quality Assessment Laboratory
Capital Management Consulting, Inc.
Carl Schaper
Checkpoint Development Inc.
Chiron Corporation
ChromaVision Medical Systems, Inc.
Chronolab Ag
Clinical Design Group Inc.
Clinical Laboratory Improvement Consultants
Cognigen
Community Medical Center (NJ)
Control Lab (Brazil)
Copan Diagnostics Inc.
Cosmetic Ingredient Review
Cubist Pharmaceuticals
Dade Behring Inc. - Deerfield, IL
Dade Behring Inc. - Glasgow, DE
Dade Behring Inc. - Marburg, Germany
Dade Behring Inc. - Sacramento, CA
Dade Behring Inc. - San Jose, CA
Diagnostic Products Corporation
Eiken Chemical Company, Ltd.
Electa Lab s.r.l.
Enterprise Analysis Corporation
Essential Therapeutics, Inc.
EXPERTech Associates, Inc.
F. Hoffman-La Roche AG
Fort Dodge Animal Health
General Hospital Vienna (Austria)
Gen-Probe
GlaxoSmithKline
Greiner Bio-One Inc.
Helena Laboratories
Home Diagnostics, Inc.
Immunicon Corporation
Instrumentation Laboratory
International Technidyne Corporation
IntraBiotics Pharmaceuticals, Inc.
I-STAT Corporation
Johnson and Johnson Pharmaceutical Research and Development, L.L.C.
Kendall Sherwood-Davis & Geck
LAB-Interlink, Inc.
Laboratory Specialists, Inc.
Labtest Diagnostica S.A.
LifeScan, Inc. (a Johnson & Johnson Company)
Lilly Research Laboratories
Macemon Consultants
Medical Device Consultants, Inc.
Merck & Company, Inc.
Minigrip/Zip-Pak
Molecular Diagnostics, Inc.
mvi Sciences (MA)
Nabi
Nichols Institute Diagnostics (Div. of Quest Diagnostics, Inc.)
Nissui Pharmaceutical Co., Ltd.
Nippon Becton Dickinson Co., Ltd.
Norfolk Associates, Inc.
Novartis Pharmaceuticals Corporation
Ortho-Clinical Diagnostics, Inc. (Raritan, NJ)
Ortho-Clinical Diagnostics, Inc. (Rochester, NY)
Oxoid Inc.
Paratek Pharmaceuticals
Pfizer Inc
Pharmacia Corporation
Powers Consulting Services
Premier Inc.
Procter & Gamble Pharmaceuticals, Inc.
The Product Development Group
QSE Consulting
Quintiles, Inc.
Radiometer America, Inc.
Radiometer Medical A/S
David G. Rhoads Associates, Inc.
Roche Diagnostics GmbH
Roche Diagnostics, Inc.
Roche Laboratories (Div. Hoffmann-La Roche Inc.)
Sarstedt, Inc.
SARL Laboratoire Carron (France)
Schering Corporation
Schleicher & Schuell, Inc.
Second Opinion
Showa Yakuhin Kako Company, Ltd.
Streck Laboratories, Inc.
SurroMed, Inc.
Sysmex Corporation (Japan)
Sysmex Corporation (Long Grove, IL)
The Clinical Microbiology Institute
The Toledo Hospital (OH)
Trek Diagnostic Systems, Inc.
Versicor, Inc.
Vetoquinol S.A.
Visible Genetics, Inc.
Vysis, Inc.
Wallac Oy
Wyeth-Ayerst
Xyletech Systems, Inc.
YD Consultant
YD Diagnostics (Seoul, Korea)

Trade Associations

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

Associate Active Members

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

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

Klinični Center (Slovenia)
Laboratories at Bonfils (CO)
Laboratoire de Santé Publique du Quebec (Canada)
Laboratório Fleury S/C Ltda. (Brazil)
Laboratory Corporation of America (NJ)
Laboratory Corporation of America (MO)
LAC and USC Healthcare Network (CA)
Lakeland Regional Medical Center (FL)
Lancaster General Hospital (PA)
Langley Air Force Base (VA)
LeBonheur Children's Medical Center (TN)
Libero Instituto Univ. Campus BioMedico (Italy)
Long Beach Memorial Medical Center (CA)
Louisiana State University Medical Center
Maccabi Medical Care and Health Fund (Israel)
Magee Womens Hospital (PA)
Malcolm Grow USAF Medical Center (MD)
Manitoba Health (Winnipeg, Canada)
Martin Luther King/Drew Medical Center (CA)
Massachusetts General Hospital (Microbiology Laboratory)
MDS Metro Laboratory Services (Burnaby, BC, Canada)
Medical College of Virginia Hospital
Medicare/Medicaid Certification, State of North Carolina
Memorial Medical Center (IL)
Memorial Medical Center (LA) Jefferson Davis Hwy
Memorial Medical Center (LA) Napoleon Avenue
Mescalero Indian Hospital (NM)
Methodist Hospital (TX)
Methodist Hospitals of Memphis (TN)
MetroHealth Medical Center (OH)
Michigan Department of Community Health
Mississippi Baptist Medical Center
Monte Tabor – Centro Italo - Brazileiro de Promocao (Brazil)
Montreal Children's Hospital (Canada)
Montreal General Hospital (Canada)
MRL Pharmaceutical Services, Inc. (VA)
MRL Reference Laboratory (CA)
Nassau County Medical Center (NY)
National Institutes of Health (MD)
Naval Surface Warfare Center (IN)
Nebraska Health System
New Britain General Hospital (CT)
New England Fertility Institute (CT)
North Carolina State Laboratory of Public Health
Northern Indiana Education Foundation
North Kansas City Hospital (MO)
North Shore – Long Island Jewish Health System Laboratories (NY)
Northwestern Memorial Hospital (IL)
O.L. Vrouwziekenhuis (Belgium)
Ordre professionnel des technologists médicaux du Québec
Ospedali Riuniti (Italy)
The Ottawa Hospital (Ottawa, ON, Canada)
Our Lady of Lourdes Hospital (NJ)
Our Lady of the Resurrection Medical Center (IL)
Pathology and Cytology Laboratories, Inc. (KY)
The Permanente Medical Group (CA)
Piedmont Hospital (GA)
Pikeville Methodist Hospital (KY)
Pocono Hospital (PA)
Presbyterian Hospital of Dallas (TX)
Queen Elizabeth Hospital (Prince Edward Island, Canada)
Queensland Health Pathology Services (Australia)
Quest Diagnostics Incorporated (CA)
Quintiles Laboratories, Ltd. (GA)
Regions Hospital
Reid Hospital & Health Care Services (IN)
Research Medical Center (MO)
Rex Healthcare (NC)
Rhode Island Department of Health Laboratories
Riyadh Armed Forces Hospital (Saudi Arabia)
Royal Columbian Hospital (New Westminster, BC, Canada)
Sacred Heart Hospital (MD)
Saint Mary's Regional Medical Center (NV)
St. Alexius Medical Center (ND)
St. Anthony Hospital (CO)
St. Anthony's Hospital (FL)
St. Barnabas Medical Center (NJ)
St-Eustache Hospital (Quebec, Canada)
St. Francis Medical Ctr. (CA)
St. John Hospital and Medical Center (MI)
St. John Regional Hospital (St. John, NB, Canada)
St. Joseph Hospital (NE)
St. Joseph's Hospital – Marshfield Clinic (WI)
St. Joseph Mercy Hospital (MI)
St. Luke's Regional Medical Center (IA)
St. Mary Medical Center (IN)
St. Mary of the Plains Hospital (TX)
St. Mary's Hospital & Medical Center (CO)
St. Paul's Hospital (Vancouver, BC, Montreal)
St. Vincent Medical Center (CA)
Ste. Justine Hospital (Montreal, PQ, Canada)
Salina Regional Health Center (KS)
San Francisco General Hospital (CA)
Santa Clara Valley Medical Center (CA)
Seoul Nat'l University Hospital (Korea)
Shanghai Center for the Clinical Laboratory (China)
South Bend Medical Foundation (IN)
Southwest Texas Methodist Hospital (TX)
South Western Area Pathology Service (Australia)
Southern Maine Medical Center
Specialty Laboratories, Inc. (CA)
Stanford Hospital and Clinics (CA)
State of Washington Department of Health
Stony Brook University Hospital (NY)
Stormont-Vail Regional Medical Center (KS)
Sun Health-Boswell Hospital (AZ)

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

Contents

Abstract i

Committee Membership v

Active Membership vii

Foreword xv

1 Introduction 1

1.1 Purpose 1
1.2 Consensus on Statistical Application 1
1.3 Literature Citations 1

2 Scope 2

3 Definitions 2

4 Path of Workflow 3

5 Phase 1: Plan 3

5.1 Define Requirements 3
5.2 Select Test System to Be Evaluated 3
5.3 Scope of the Evaluation Study 3
5.4 Selection of the Study Protocol 6

6 Phase 2: Perform Testing or “Do” 6

6.1 Confidence Intervals Versus Hypothesis Testing 6
6.2 Hypothesis Tests 8
6.3 Confidence Intervals 8

7 Phase 3: Check (and Communicate Conclusion) 10

8 Phase 4: Implement or Act 11

9 Conclusion 11

References 12

Related NCCLS Publications 13

Foreword

This report presents an overview of the relationships among evaluation experiments that are performed by manufacturers, by clinical laboratories, and/or that may be required by governmental agencies. It is the objective of the NCCLS Area Committee on Evaluation Protocols to provide recommendations for the assessment of laboratory test procedures in a way that defines good laboratory practices that are grounded on good scientific and statistical theory.

NCCLS evaluation protocols are primarily intended to be practical in their use in the clinical laboratory. All clinical laboratories must perform method evaluation studies prior to using a new or revised test system for reporting patient test results. The extent of a given study may vary considerably depending on the objectives of the laboratory, on resources available in the laboratory, and on information that may have been previously reported in technical literature by other laboratories.

NCCLS evaluation protocols are not intended to be so comprehensive as to provide the sole data on product testing and qualification on the part of manufacturers of *in vitro* diagnostic devices. There are many other studies that a manufacturer will perform during the research and development phases and during the manufacturing validation phase that are unique to the design of the test kit or test system and the manufacturing processes themselves. These special studies go beyond the scope of NCCLS evaluation protocols. However, it is appropriate that manufacturers use these NCCLS documents as a means of generating performance data for the final product in a form to which the laboratory can relate.

In addition to defining a framework for a family of evaluation studies that should result in greater consensus and standardization of practice, this report reviews two different approaches for statistical analysis of the results: tests of significance and the use of confidence interval estimates. The latter is preferred, so that the laboratorian focuses attention on the magnitude of an error assessment in the comparison to stated claims or stated requirements from a clinical perspective.

EP19 was originally conceived as a report and was primarily intended for use by evaluation protocols subcommittees. However, as draft copies were circulated during its preparation, some NCCLS users outside of the evaluation protocols area committee who reviewed the report found intrinsically valuable information in the document and suggested that it be made available as a guideline. The Area Committee on Evaluation Protocols concurred with users' recommendations and issued EP19 as a proposed-level consensus document. However, due to the limited input received on EP19-P during the public review and comment period, the area committee agreed that EP19 be reissued as a report, made available to all NCCLS users.

Key Words

Demonstration, evaluation protocol, validation, verification

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 document HS1—*A Quality System Model for Health Care.* The quality system approach applies a core set of "quality system essentials (QSEs)," basic to any organization, to all operations in any healthcare service's path of workflow. The QSEs provide the framework for delivery of any type of product or service, serving as a manager's guide. The quality system essentials (QSEs) are:

QSEs

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

EP19-R 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.*

A Framework for NCCLS Evaluation Protocols; A Report

1 Introduction

It is the objective of the NCCLS Area Committee on Evaluation Protocols to provide recommendations for the assessment of laboratory test procedures in a way that defines good laboratory practices that are grounded on good science and statistical theory. These recommendations are primarily intended to be practical in their use in the clinical laboratory. They are not intended to be so comprehensive as to be the sole data on product testing and qualification on the part of manufacturers of *in vitro* diagnostic devices. However, it is appropriate that manufacturers use these documents as a means of generating performance data to which the laboratory can relate.

1.1 Purpose

It is recognized that laboratories may have different needs and purposes for conducting evaluation studies; they may have different resources to apply to an evaluation study; and there may be different performance requirements.

This report creates a framework that relates the scope and design of an evaluation study to the purpose of the study. The framework also integrates this hierarchy of evaluation studies with the amount of performance information that may be known prior to the study. (This is discussed in greater detail in Section 2, Scope.)

Finally, this report incorporates the fundamentals of a defined path of workflow[1-3] for an evaluation study as described in Section 4. Please refer to the most recent edition of NCCLS document HS1—*A Quality System Model for Health Care* for additional information. All evaluation studies, regardless of purpose and scope have certain key elements or steps in common. Key issues that have been sources of confusion are discussed in more detail in the appropriate steps in the overall process for an evaluation study.

1.2 Consensus on Statistical Application

Implied in the discussion of purpose is the requirement that there be a common understanding of the best approach in the application of statistics for data analysis and interpretation. Section 6.1 states that the use of confidence interval estimates of error is preferred over tests of statistical significance.[4-10] Confidence interval estimates reflect both the magnitude of an error as well as the uncertainty of that estimate, while tests of significance are based on a ratio of errors, so that the magnitude of the error may be lost in the final interpretation. It is recognized that in the clinical laboratory setting, the magnitude of the error due to a given effect is of greater importance than whether the error is "statistically" significant. The ultimate decision regarding the acceptability of the performance of a given test or product should be based on whether the magnitude of a given error is less than the medical or regulatory requirement.

1.3 Literature Citations

It is recommended that statistical methods, whether for experimental design, data analysis, or interpretation of data, be referenced from peer-reviewed journals and/or established reference textbooks. Several NCCLS guidelines have been published with little or no reference to the technical literature, and this should be corrected. New documents and those under development should be adequately referenced prior to the next stage in the consensus process. Existing documents should have appropriate references added at the time of their next review.

2 Scope

This report is intended for use by all constituent groups of NCCLS—clinical laboratory professionals, manufacturers, and governmental agencies—as a basis for arriving at a consensus for the need for different evaluation studies that are designed to meet different purposes.

NCCLS evaluation protocols are not intended to be so comprehensive as to provide the sole data on product testing and qualification on the part of manufacturers of *in vitro* diagnostic devices. There are many other studies that a manufacturer will perform during the research and development phases and during the manufacturing validation phase that are unique to the design of the test kit or test system and the manufacturing processes themselves. These special studies go beyond the scope of NCCLS evaluation protocols. However, it is appropriate that manufacturers use these recommendations as a means of generating performance data for the final product in a form to which the laboratory can relate.

3 Definitions[a]

Alpha level, *n* – The probability that one will reject the null hypothesis when it is true; **NOTE:** The alpha level is often set at 5 or 1 percent.

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

Demonstration (study), *n* – A study performed by a laboratory to show that it is capable of using a test system to obtain expected performance.

Evaluation (study), *n* – A generic term for any study that measures the performance capabilities of an assay.

Hypothesis testing, *n* – ***In Statistics,*** the testing of two or more statistical hypotheses that are mutually exclusive so that exactly one hypothesis can be accepted at a specified confidence level.

Null hypothesis, *n* – The hypothesis of no difference or no differential effects beyond random variation.

Point estimate, *n* – A value that summarizes a set of data without accounting for the precision of the estimate (e.g., its uncertainty).

Validation, *n* – **1)** Confirmation, through the provision of objective evidence, that requirements for a specific intended use or application have been fulfilled; **NOTES:** a) The term "validated" is used to designate the corresponding status (*ISO 9000:2000*); b) The use conditions for validation can be real or simulated; **2)** A term used by the FDA for a study used to determine whether a test system meets user needs. ***(Federal Register: Part VII Dept. of Health and Human Services, FDA. 12 CFR Parts 808, 812, and 820. Rules and Regulations. October 7, 1996;61(195):52222–52601)***; **3)** The action {or process} of proving that a procedure, process, system, equipment, or method used works as expected and achieves the intended result.

Verification, *n* – **1)** Confirmation, through the provision of objective evidence, that specified requirements have been fulfilled; **NOTES:** a) The term "verified" is used to designate the corresponding status; b) Confirmation can comprise activities such as: performing alternative calculations; comparing a

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

new design specification with a similar proven design specification; undertaking tests and demonstrations; and reviewing document prior to issue (*ISO 9000:2000*); **2)** A term used by the FDA for a study used to determine whether a test system meets specifications. ***(Federal Register: Part VII Dept. of Health and Human Services, FDA. 21 CFR Parts 808, 812, and 820. Rules and Regulations. October 7, 1996;61(195):52222–52601).***

4 Path of Workflow

The process for an evaluation study can be defined in terms of the four steps of the Deming quality improvement cycle: Plan, Do, Check, Act. In general application of this process for continuous quality improvement, the cycle is continuously repeated. However, in the situation where a method evaluation study is performed, the evaluation study is usually performed one time, prior to the use of a given procedure or test system for patient testing. In this way, the laboratory management and staff can know that the test system will meet their needs and the needs of their medical staff before it is used to test patient samples.

5 Phase 1: Plan

For a given study, it is essential that certain details be planned out to ensure that the study is sufficiently complete so that it has to be done only one time. The performance requirements should be defined before the evaluation has been initiated. These goals might be based on the manufacturer's claims on proficiency testing requirements, and on medical requirements defined by the clinicians who use the laboratory services. The scope of the study will impact on resources required. The type of test system will impact on training required, on technical expertise, and on the amount of "practice" required to ensure that the results reflect what will be obtained under routine use.

5.1 Define Requirements

The following requirements should be defined:

- managerial requirements, including cost, size, maintenance, sample size, etc.; and
- performance requirements, including total error, precision, bias, linearity, etc.

5.2 Select Test System to Be Evaluated

Selection of the test system for evaluation should be based on the requirements that have been defined. Any system that is to be considered a candidate must meet the practical requirements. The performance requirements may be used to select among the possible candidates.

5.3 Scope of the Evaluation Study

In order to select a protocol, the scope of the evaluation will be dependent on the purpose of the study. To differentiate among purposes, this report recommends the following.

5.3.1 Evaluate

This report recommends the use of the word "evaluate" as a generic term for any study that measures the performance capabilities of an assay. Several subclasses of evaluation studies are defined below. For the last three terms used below, "validate" is the most comprehensive, followed by "verify," and "demonstrate." "Establishment" studies are defined but not discussed further.

Evaluation protocols include:

- establishment studies (R&D);
- validation protocols;
- verification protocols; and
- demonstration protocols.

5.3.2 Establish Performance

The type of studies performed to "establish performance" are those that are performed by a manufacturer or research scientist in the research and development (R&D) or manufacturing phases to enable the manufacturer to define the parameters of the components and processes to achieve the intended product design. These are considered beyond the scope of NCCLS evaluation protocols, since in most cases, they are specific to the product design and manufacturing processes.

5.3.3 Validate

This term refers to evaluation studies that are performed to determine whether a test system meets defined user needs, where "user" is taken to mean the laboratory. This testing is performed either at a manufacturer or in an end-use laboratory. The term "external validation" is used if the evaluation is performed in an end user's environment. To validate a performance characteristic relative to user requirements, a statistically valid sample size must be included in the study design.

As alluded to in the introduction, the scope of a study may also be impacted by what is already known about the performance of a test system. For a new test system that has just been created by a manufacturer, it is necessary for the manufacturer to perform internal validation studies. After these have been shown to be successful, then the test system should be evaluated by an external validation by one or more selected end-user laboratories to see if the test system meets user requirements. This information may become part of the information used by the manufacturer to define the claims for performance. Once these claims have been defined, and the test system is available on the market for clinical use, then other laboratories may wish to perform external validation studies to confirm that the test system meets the claimed specifications.

5.3.4 Verify

This term refers to evaluation studies that are performed to determine whether certain specifications are met.[11] A manufacturer performs internal verification studies to confirm that the test system meets internal specifications that were based on laboratory requirements. A laboratory may perform verification studies to confirm that the test systems meet the specifications claimed by the manufacturer in the product literature (or product insert sheets). These verification studies might be referred to as "external verification studies." To verify a performance characteristic, a statistically valid sample size must be included in the study design.

After a number of these types of external studies (external verification or external validation) have been completed and published in the literature or by other communication mechanisms, one might conclude that the performance of such a test system is fairly well-known, and there is no need to validate or verify performance again. Nevertheless, it is incumbent for each laboratory to document that when such a test system is introduced into the laboratory, the laboratory staff is able to obtain "expected performance." Expected performance might be that based on performance documented elsewhere, such as by proficiency testing.

The distinction between verification and validation is shown in Figure 1, which describes a device design, development, and implementation process.

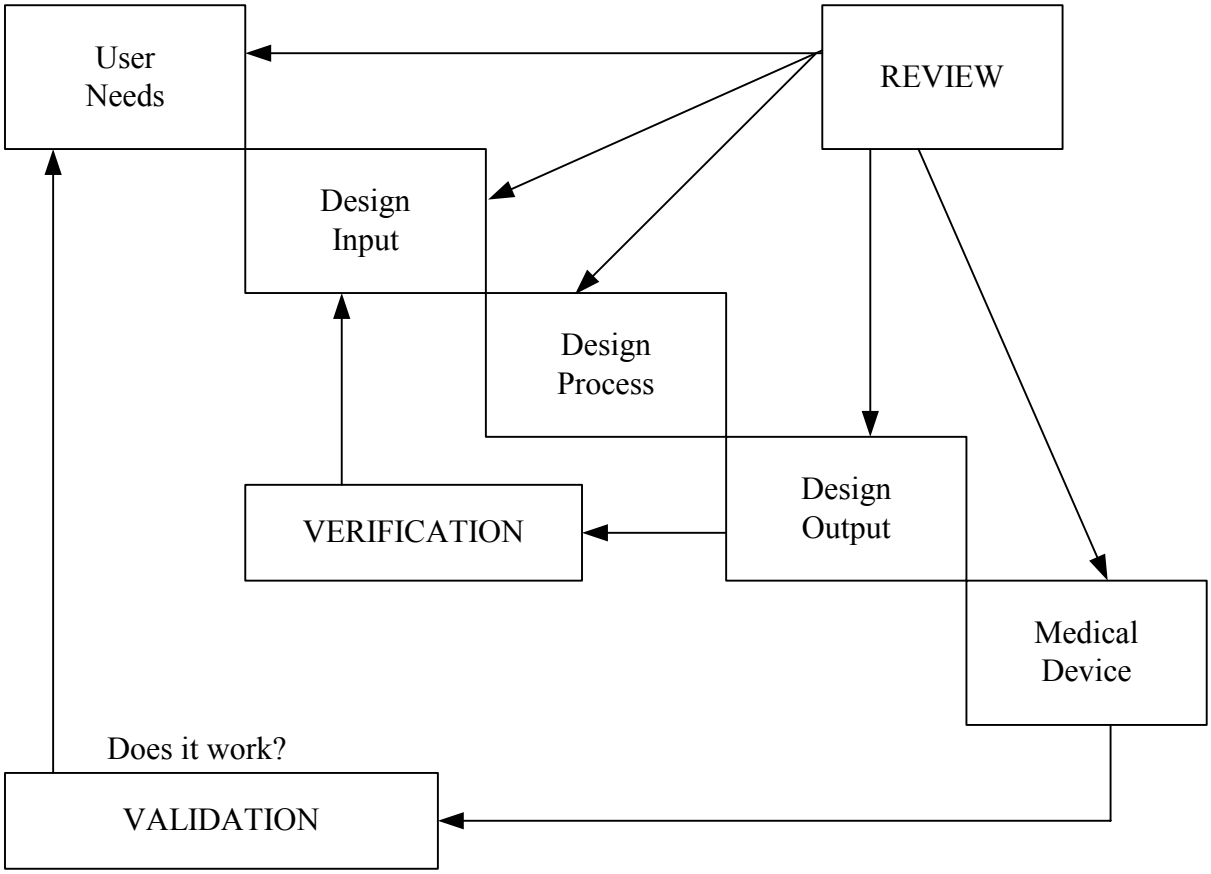


Figure 1. Design Acceptance. Adapted from the Federal Register: Part VII Dept. of Health and Human Services, FDA. 21 CFR Parts 808, 812, and 820. Rules and Regulations. October 7, 1996;61(195):52222–52601.

5.3.5 Demonstrate

This term refers to an evaluation study that is performed by a laboratory to enable them to show (demonstrate) that it is capable of using the test system to obtain expected performance. Examples of when demonstration evaluation would be performed include:

- **Demonstration of performance prior to implementation of a new test.**
- **Demonstration of performance after failure of proficiency testing.** In the case where a laboratory has failed proficiency testing, and after implementing suitable corrective actions, the laboratory can use this type of protocol to show that acceptable analytical performance is now being obtained.
- **Demonstration of performance to document technical competency of testing personnel.** This protocol may be used in the situation where laboratory staff uses this type of protocol to show abilities relative to the annual technical competency assessment.

Table 1 below shows a relationship between the study design and the amount of information that is known prior to performing such a study.

Table 1. Relationship Between Study Design and Information Known Prior to Study

Item to Test	Study
Design specifications	Verification
User requirements	Validation (Internal or External)
User requirements for specific laboratory	Demonstration

5.4 Selection of the Study Protocol

Once the scope of the evaluation study has been defined, the laboratory manager can select from appropriate NCCLS protocols to define the actual day-to-day tasks to be performed in the study. The laboratory staff should review this and perform several practice runs to ensure their awareness of the details of the study.

6 Phase 2: Perform Testing or "Do"

The actual testing details are described in the appropriate NCCLS evaluation protocol. Current protocols may not address all purposes and scopes as discussed above. Table 2 shows a summary of how current protocols may be related to the framework described in this document. Please refer to the appropriate document for more details.

Table 2. Current NCCLS Evaluation Protocols

Guideline	Title	Purpose*
EP5	Evaluation of Precision Performance	Val, Ver
EP6	Evaluation of the Linearity of Quantitative Analytical Methods	Val, Ver, Demo
EP7	Interference Testing	Val, Ver
EP9	Method Comparison and Bias Estimation Using Patient Samples	Val, Ver
EP10	Preliminary Evaluation of Quantitative Clinical Laboratory Methods	Demo
EP12	Evaluation of Qualitative Test Performance	Val, Ver, Demo
EP14	Evaluation of Matrix Effects	Val, Ver
EP15	User Demonstration of Performance for Precision and Accuracy	Demo
EP18	Unit-Use Testing	N/A
EP21	Total Error for Clinical Laboratory Methods	Val, Ver, Demo

*Val = "validate," Ver = "verify," and Demo = "demonstrate."

A key element in any evaluation is the application of statistics, both in terms of study design and in terms of result summarization to facilitate interpretation. Applications of statistics in study design are incorporated into each protocol. Applications of statistical techniques for proper interpretation of results are discussed in the next section.

6.1 Confidence Intervals Versus Hypothesis Testing

All scientific investigations have, or should have, a goal. Often, a study involves measurements that exhibit variation. With variation present, a statistically planned experiment is required to answer the question raised by the goal. Figure 2 shows a flow chart of the experimentation process. Table 3 shows that different goals lead to different data analysis methods. Thus, a goal can be either to answer a question with a yes/no response (e.g., verify) or to characterize something by estimating its value. For example, one can ask whether a new cholesterol method versus the reference method has a slope detectably different from one. Statistically, this goal results in use of a hypothesis test. Alternatively, one

can ask what is the actual slope of the new cholesterol method versus the reference method. Statistically, this results in estimating the slope and its confidence interval.

The purpose of this section is to compare the salient features of these two statistical procedures. The focus shall be on information content, ease of use, reliability, and interchangeability. This document is not a tutorial on how to perform these methods.

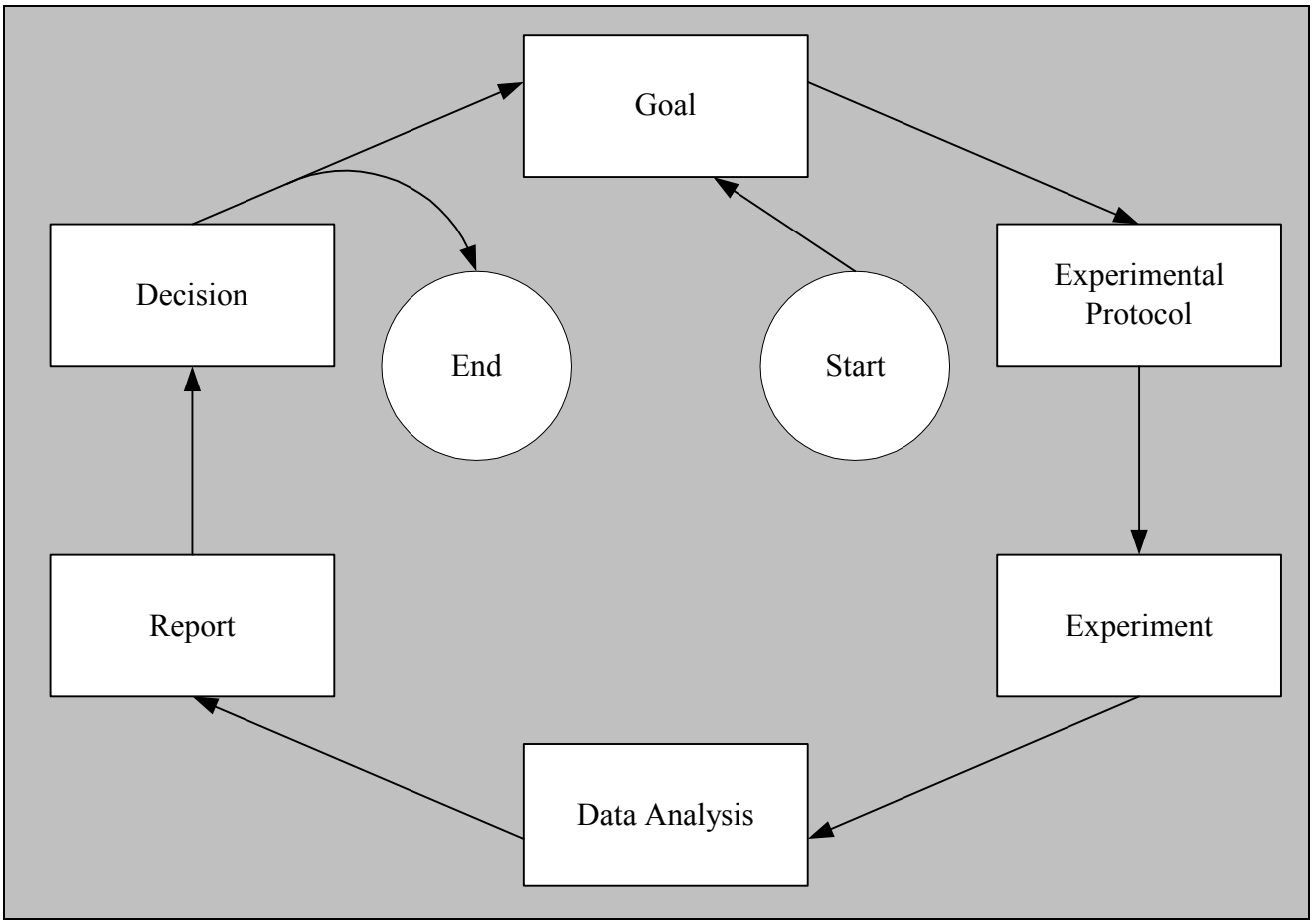


Figure 2. Flow Chart of the Experimental Process

Table 3. How Different Goals are Achieved Statistically

Goal	Data Analysis Method	Examples
Verify (yes/no)	Hypothesis Test	t-test F-test
Characterize	Estimation	Point Estimate Confidence Interval Prediction Interval Tolerance Interval

6.2 Hypothesis Tests

Hypothesis tests entail the following steps:

- State the null hypothesis.
  - — state the alternative hypothesis
  - — calculate the sample size
- Estimate the parameter.
- Perform the test.

The steps indented with a dash are often omitted, largely because they are difficult for nonstatisticians. Thus, the ease of use of the complete hypothesis test is low. The information content is also low, because one has dichotomized the result of a continuous variable. For example, using a hypothesis test one would answer "yes" to the question, "Is the patient hyperglycemic?" regardless of whether the glucose value was 150 or 500 mg/dL. People informally recognize this shortcoming of hypothesis testing by trying to attach different significance to different p values, such as "P = 0.03 is borderline" and "P less than 0.001 is highly significant." The problem with this approach is it is difficult to interpret values of P closer to each other (e.g., P = 0.002 vs. P = 0.009). Thus, the P scale, in spite of being continuous, is mainly interpreted as a categorical scale, often with only the two states, significant or not significant.

Reliability of the result will be high if all steps are followed and undetermined if the indented steps are omitted. This latter point can be shown by the following example. The question asked is: Does bilirubin interfere in a cholesterol assay? The hypothesis test procedure above is followed whereby the indented steps are skipped. In this example, the null hypothesis is "no interference" and the alpha level is 0.05. Assume the test result is nonsignificant (P greater than 0.05). The result is "no bilirubin interference" or, more correctly, "no bilirubin interference detected." Yet, assume that a 95% confidence interval was also calculated, and it was found that it covered –23 to 25 mg/dL. This means that bilirubin interference could be as high as 25 mg/dL. Hence, the reliability of the hypothesis test, with the indented steps of the procedure omitted, is low. Specifically, the possibility of false-negative results is left open.

6.3 Confidence Intervals

Confidence intervals shall be discussed; note that the discussion also applies to prediction and tolerance intervals. Confidence intervals entail the following steps:

- — State either the desired length of the confidence interval or deviation from target that one does not wish to exceed.
- State the alpha level.
  - — Calculate the sample size.
- Estimate the parameter.
- Calculate the confidence interval.

Once again, the intended steps are not always followed. Yet, even if one does not specify the deviation from target that one does not wish to exceed, the calculation of the confidence interval gives this

information, albeit respectively. Ease of use is similar to hypothesis testing, since in both procedures steps are omitted.

An advantage of using a confidence interval is that the parameter estimate and interval are on a continuous scale in the original units which provides more information than a yes/no response. Moreover, the length of the confidence interval affords a measure of the result's reliability. Reliability and clinical importance can be judged simultaneously. With either a hypothesis test or confidence interval, one needs to estimate the parameter (the point estimate). However, with a hypothesis test this estimate is incidental to the hypothesis test, whereas it is a key part of the confidence interval.

In many cases one can perform both procedures using data from the same experiment, but it is not always clear that this is meaningful. For example, to assess linearity, one can use a hypothesis test (lack of fit) to detect the presence of nonlinearity. There is no commonly performed confidence interval procedure that corresponds to this. On the other hand, one can estimate with a confidence interval the magnitude due to nonlinearity at a specific concentration level. In a sense, different questions are being answered (e.g., Is there global nonlinearity vs. the magnitude of nonlinearity at a specific concentration level?). Considering these questions up front is recommended as a means of deciding which statistical method to select.

For both procedures, discussing the necessary steps common to either method, checking the data for outliers, assuming a certain distribution, or any other step required by the estimation procedure have been omitted. This leads to a caution. With the additional information that is provided by a confidence interval, one may feel more secure in making decisions. Yet, a confidence interval will be accurate only if all required assumptions are true. The most important assumption and the one that is the most difficult to verify is that "the data are a representative sample from the population about which interferences are desired." For example, if one wishes to make a decision about the quality of results of an assay, one would perform an evaluation. If the assay is run on a random access analyzer routinely, and one performs the evaluation in a batch mode, the data will not be a representative sample. A famous example involves estimation of the solar unit by well-known scientists over a one-hundred-year period. Although each estimate had a confidence interval, each scientist's confidence interval failed to overlap the confidence interval of his predecessor.[12]

Most scientists recognize the importance of the assumption of representative data by using judgment in interpreting results and are thus aware of the syndrome *statistics on, brain off*.

Finally, it is recommended to display results graphically where possible. An example of a graphical display of confidence intervals for interference studies is shown below (Figure 3), taken from the most current version of NCCLS document EP7—*Interference Testing in Clinical Chemistry*.

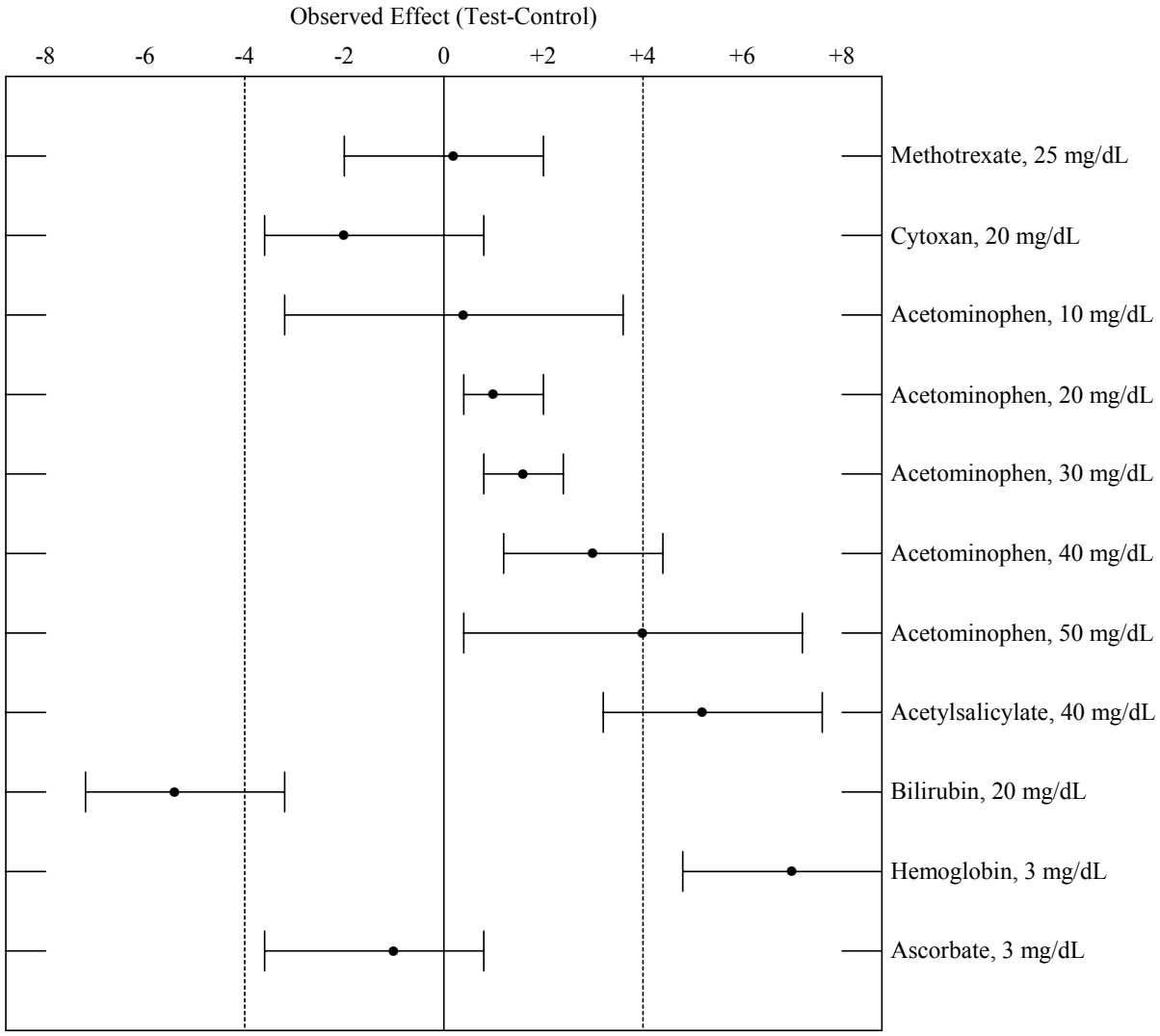


Figure 3. Graphical Display of Confidence Intervals

7 Phase 3: Check (and Communicate Conclusion)

All evaluation studies require some type of decision-making step, either to compare to the manufacturer's claims for an external verification study, or to compare to user requirements for an external validation study. In a demonstration study, one can compare to either the manufacturer's claims or to other sources of expected performance. This step should be described explicitly in each evaluation guideline or protocol.

Once the assessment of performance has been made, then the conclusion must be communicated. There are many parties that need to know the conclusion:

- laboratory staff;
- purchasing department;
- manufacturer; and
- medical staff

to name a few, and perhaps in that order.

8 Phase 4: Implement or Act

The actual evaluation process is not completed until the new test system has been implemented. This is the sole reason for performing the studies in the first place, except perhaps for those few academic research-oriented evaluations.

The implementation phase should describe:

- establishment of supply process;
- training of other laboratory staff;
- definition of the maintenance procedure;
- definition and establishment of the QC program (to monitor implementation);
- integration to the LIS system;
- notification to the medical staff; and
- feedback process from the medical staff.

9 Conclusion

It is essential that the clinical laboratory community (manufacturer, professional, and government) speak a common language with regard to the whole process of an evaluation study, which entails

- planning;
- performing the studies;
- analyzing the data and interpreting the results;
- reporting the results, conclusions, and recommendations;
- implementing the new test system;
- the different objectives for evaluations and how these lead to studies of different scope and design; and
- the appropriate use of statistics for data analysis and interpretation.

References

[1] Garber CC, Carey RN. "Evaluation of Methods." In: Kaplan LA, Pesce JP, eds. *Clinical Chemistry: Theory Analysis, Correlation.* 3rd ed. St. Louis, MO: C.V. Mosby Co.; 1996;22:402-423.

[2] Koch DD, Peters T Jr. "Selection and Evaluation of Methods." In: Burtis CA, Ashwood ER, eds. *Tietz Textbook of Clinical Chemistry*. 3rd ed. Philadelphia, PA: W.B. Saunders; 1999;13:321-335.

[3] Westgard JO. *Basic Method Validation*. Madison, WI: Westgard Quality Corporation; 1999.

[4] Scialli AR. Confidence and the null hypothesis. *Reprod Toxicol*. 1992;6(5):383-384.

[5] Birnbaum D, Sheps SB. The merits of confidence intervals relative to hypothesis testing. *Infect Control Hosp Epidemiol*. 13:553-555.

[6] Johnson LA. Analysis of variance of parameter estimates: F tests and t tests. *Anal Biochem*. 1992;206(1):195-201.

[7] Harris EK. On P values and confidence intervals. *Clin Chem*. 1993;39:927-928.

[8] Henderson AR. Chemistry with confidence: should clinical chemistry require confidence intervals for analytical and other data? *Clin Chem*. 1993;39:929-935.

[9] Selvin S, White MD. Description and reporting of statistical methods. *Am J Infect Control.* 1993;21:210-215.

[10] Altman DG. Confidence intervals in clinical chemistry. *Clin Chem*. 1994;40:161-162.

[11] Federal Register: Part VII Dept. of Health and Human Services, FDA. 21 CFR Parts 808, 812, and 820. Rules and Regulations. October 7, 1996;61(195):52222–52601.

[12] Youden WJ. Enduring values. *Technometrics.* 1972;14:1-11.

Related NCCLS Publications*

EP5-A **Evaluation of Precision Performance of Clinical Chemistry Devices; Approved Guideline (1999).** This document provides guidance for designing an experiment to evaluate the precision performance of clinical chemistry devices; recommendations on comparing the resulting precision estimates with manufacturer's precision performance claims and determining when such comparisons are valid; as well as manufacturer's guidelines for establishing claims.

EP6-P2 **Evaluation of the Linearity of Quantitative Analytical Methods; Proposed Guideline—Second Edition (2001).** This document provides guidelines for characterizing the linearity of a method during a method evaluation; for checking linearity as part of routine quality assurance; and for determining and stating a manufacturer's claim for linear range.

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

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

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

EP12-P **User Protocol for Evaluation of Qualitative Test Performance; Proposed Guideline (2000).** This document contains guidelines that optimize the experimental design for the evaluation of qualitative tests to better measure performance and provide a structured 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 results from the sample matrix (physiological or artificial) when two analytical methods are compared.

EP15-A **User Demonstration of Performance for Precision and Accuracy; Approved Guideline (2001).** This document demonstrates method precision and accuracy for laboratory analyte determinations, utilizing a protocol designed to be completed within five or fewer working days.

EP18-P **Quality Management for Unit-Use Testing; Proposed Guideline (1999).** This document recommends a quality management system for unit-use devices that will aid in the identification, understanding, and management of sources of error and help to ensure correct results. It is targeted for those involved in the supervision of laboratory-testing quality management, and it addresses issues related to specimen collection through reporting of test results.

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

Related NCCLS Publications (Continued)

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).

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 1-56238-464-3