EP15-A
Vol. 21 No. 25

Replaces EP15-P
Vol. 18 No. 22

User Demonstration of Performance for Precision and Accuracy; Approved Guideline

This document describes the demonstration of method precision and accuracy for laboratory analyte determinations utilizing a protocol designed to be completed within five working days or less.

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

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.

User Demonstration of Performance for Precision and Accuracy; Approved Guideline

Abstract

NCCLS document EP15-A—*User Demonstration of Performance for Precision and Accuracy; Approved Guideline* describes the demonstration of method precision and accuracy for analyte determinations performed within the laboratory. Included are guidelines for the duration, procedures, materials, data summaries, and interpretation techniques that are adaptable for the widest possible range of analytes and device complexity. A balance is created in the document between the complexity of design and formulae, and the simplicity of operation. The protocol is designed to be completed within five working days or less. Definitions are provided for "within-run" and "total" precision.

NCCLS. *User Demonstration of Performance for Precision and Accuracy; Approved Guideline*. NCCLS document EP15-A (ISBN 1-56238-451-1). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2001.

EP15-A
ISBN 1-56238-451-1
ISSN 0273-3099

User Demonstration of Performance for Precision and Accuracy; Approved Guideline

Volume 21 Number 25

R. Neill Carey, Ph.D.
Harvey George, Ph.D.
Alfred E. Hartmann, M.D.
Verlin K. Janzen, M.D.
Jacob B. Levine, M.B.A.
W. Gregory Miller, Ph.D.
Aparna Srinivasan, Ph.D.

Suggested Citation

(NCCLS. *User Demonstration of Performance for Precision and Accuracy; Approved Guideline.* NCCLS document EP15-A [ISBN 1-56238-451-1]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2001.)

Proposed Guideline
December 1998

Approved Guideline
December 2001

ISBN 1-56238-451-1
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**

Subcommittee on User Demonstration of Performance

R. Neill Carey, Ph.D. **Chairholder**	**Peninsula Regional Medical Center** **Salisbury, Maryland**
Harvey George, Ph.D.	Massachusetts Department of Public Health Labs. Jamaica Plain, Massachusetts
Alfred E. Hartmann, M.D.	McKennan Laboratories Sioux Falls, South Dakota
Verlin K. Janzen, M.D.	Hutchinson Clinic, P.A. Hutchinson, Kansas
Jacob B. Levine, M.B.A.	Bayer Diagnostics Tarrytown, New York
W. Gregory Miller, Ph.D.	Virginia Commonwealth University Richmond, Virginia
Aparna Srinivasan, Ph.D.	LifeScan Milpitas, California

Advisors

Kaiser Aziz, Ph.D.	FDA Centers for Devices/Radiological Health Rockville, Maryland
Dr. José Abol Correa	Sociedade Brasileira de Análises Clínicas Rio de Janeiro, Brazil
Kevin D. Fallon, Ph.D.	Instrumentation Laboratory Lexington, Massachusetts
William Gilbert, Ph.D.	Sigma Diagnostics St. Louis, Missouri
Gary A. Graham, Ph.D.	Rochester, New York
J. David Lyle, M.Ed., M.T.(ASCP)	FDA Centers for Devices/Radiological Health Rockville, Maryland

Advisors (Continued)

Ronald H. Ng, Ph.D.	Abbott Laboratories Bedford, Massachusetts
John A. O'Malley	Second Opinion Menlo Park, California
Donald R. Parker, Ph.D.	Bayer Corporation Elkhart, Indiana
Dr. Humberto Marques Tibúrcio	Sociedade Brasileira de Análises Clínicas Rio de Janeiro, Brazil
Jay Valinsky, Ph.D.	The New York Blood Center New York, New York
Jack Zakowski, Ph.D.	Beckman Coulter, Inc. Brea, California
Jennifer K. McGeary, M.T.(ASCP), M.S.H.A *Staff Liaison*	NCCLS Wayne, Pennsylvania
David E. Sterry, M.T.(ASCP) *Project Manager*	NCCLS Wayne, Pennsylvania
Patrice E. Polgar *Editor*	NCCLS Wayne, Pennsylvania
Donna M. Wilhelm *Assistant Editor*	NCCLS Wayne, Pennsylvania

Active Membership
(as of 1 October 2001)

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.
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 Society of Parasitologists, Inc.
American Type Culture Collection, Inc.
Asociacion de Laboratorios de Alta Complejidad
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.
Australasian Association of Clinical Biochemists
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
Canadian Society of Clinical Chemists
Clinical Laboratory Management Association
COLA
College of American Pathologists
College of Medical Laboratory Technologists of Ontario
College of Physicians and Surgeons of Saskatchewan
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
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
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 Institute of Standards and Technology
Ohio Department of Health
Ontario Ministry of Health
Saskatchewan Health-Provincial Laboratory
Scientific Institute of Public Health; Belgium Ministry of Social Affairs, Public Health and the Environment
South African Institute for Medical Research
Swedish Institute for Infectious Disease Control
Thailand Department of Medical Sciences

Industry Members

AB Biodisk
Abbott Laboratories
Abbott Laboratories, MediSense Products
Accumetrics, Inc.
Agilent Technologies, Inc.
Ammirati Regulatory Consulting
Anaerobe Systems
Asséssor
AstraZeneca
Aventis
Avocet Medical, Inc.
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, Inc.
Biometrology Consultants
Bio-Rad Laboratories, Inc.
Bio-Rad Laboratories, Inc. - France
Biotest AG
Bristol-Myers Squibb Company
Canadian External Quality Assessment Laboratory
Capital Management Consulting, Inc.
Checkpoint Development Inc.
Clinical Design Group Inc.
Clinical Laboratory Improvement Consultants
COBE Laboratories, Inc.
Community Medical Center (NJ)
Control Lab (Brazil)
Copan Diagnostics Inc.
Cosmetic Ingredient Review
Cubist Pharmaceuticals
Cytometrics, Inc.
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
DAKO A/S
Diagnostic Products Corporation
Eiken Chemical Company, Ltd.
Enterprise Analysis Corporation
Fort Dodge Animal Health
General Hospital Vienna (Austria)
Gen-Probe
GlaxoSmithKline
Greiner Bio-One Inc.
Health Systems Concepts, Inc.
Helena Laboratories
Home Diagnostics, Inc.
Immunicon Corporation
I-STAT Corporation
International Technidyne Corporation
Kendall Sherwood-Davis & Geck
LAB-Interlink, Inc.
Labtest Diagnostica S.A.
LifeScan, Inc. (a Johnson & Johnson Company)
Lilly Research Laboratories
Medical Device Consultants, Inc.
Medtronic, Inc.
Merck & Company, Inc.
mvi Sciences (MA)
Nabi
Neometrics is.
Nichols Institute Diagnostics (Div. of Quest Diagnostics, Inc.)
Nissui Pharmaceutical Co., Ltd.
Nippon Becton Dickinson Co., Ltd.
Norfolk Associates, Inc.
Organon Teknika Corporation
Ortho-Clinical Diagnostics, Inc. (Raritan, NJ)
Ortho-Clinical Diagnostics, Inc. (Rochester, NY)
Oxoid Inc.
Pfizer Inc
Pharmacia Corporation
Powers Consulting Services
Premier Inc.
Procter & Gamble Pharmaceuticals, Inc.
The Product Development Group
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.)
The R.W. Johnson Pharmaceutical Research Institute
Sarstedt, Inc.
Carl Schaper
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 Toledo Hospital (OH)
Trek Diagnostic Systems, 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)
67th CSH Wuerzburg, GE (NY)
Academisch Ziekenhuis-VUB (Belgium)
Acadiana Medical Laboratories, LTD (LA)
Adena Regional Medical Center (OH)
Advocate Laboratories (IL)
The Aga Khan Hospital & Medical College, Karachi (Pakistan)
Akershus Central Hospital and AFA (Norway)
Albany Medical Center Hospital (NY)
Albemarle Hospital (NC)
Allegheny General Hospital (PA)
Allegheny University of the Health Sciences (PA)
Allina Laboratories (MN)
Alton Ochsner Medical Foundation (LA)
American Medical Laboratories (VA)
Arkansas Department of Health
ARUP Laboratories (UT)
ARUP at University Hospital (UT)
Armed Forces Research Institute of Medical Science (APO, AP)
Aurora Consolidated Laboratories (WI)
Azienda Ospedale Di Lecco (Italy)
Bay Medical Center (MI)
Baystate Medical Center (MA)

Bbaguas Duzen Laboratories (Turkey)
Bo Ali Hospital (Iran)
Bonnyville Health Center (Alberta, Canada)
British Columbia Cancer Agency (Vancouver, BC, Canada)
Broward General Medical Center (FL)
Calgary Laboratory Services
Carilion Consolidated Laboratory (VA)
Carolinas Medical Center (NC)
Cathay General Hospital (Taiwan)
CB Healthcare Complex (Sydney, NS, Canada)
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)
Children's Hospital (LA)
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)
Clarian Health–Methodist Hospital (IN)
Clendo Lab (Puerto Rico)
CLSI Laboratories (PA)
Commonwealth of Kentucky
CompuNet Clinical Laboratories (OH)
Covance Central Laboratory Services (IN)
Danville Regional Medical Center (VA)
DeKalb Medical Center (GA)
Delaware Public Health Laboratory
Department of Health & Community Services (New Brunswick, Canada)
DesPeres Hospital (MO)
Detroit Health Department (MI)
Diagnostic Laboratory Services, Inc. (HI)
Duke University Medical Center (NC)
Durham Regional Hospital (NC)
Dynacare Laboratories - Eastern Region (Ottawa, ON, Canada)
Dynacare Memorial Hermann Laboratory Services (TX)
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)
Fairfax Hospital (VA)
Fairview-University 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)
Franklin Square Hospital Center (MD)
Fresenius Medical Care/Spectra East (NJ)
Fresno Community Hospital and Medical Center
Frye Regional Medical Center (NC)
Gambro Healthcare Laboratory Services (FL)
GDS Technology, Inc (IN)
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)
Health Sciences Centre (Winnipeg, MB, Canada)
Heartland Health System (MO)
Highlands Regional Medical Center (FL)
Hoag Memorial Hospital Presbyterian (CA)
Holmes Regional Medical Center (FL)
Holy Spirit Hospital (PA)
Holzer Medical Center (OH)
Hospital for Sick Children (Toronto, ON, Canada)
Hospital Israelita Albert Einstein (Brazil)
Hospital Sousa Martins (Portugal)
Hotel Dieu Hospital (Windsor, ON, Canada)
Huddinge University Hospital (Sweden)
Hurley Medical Center (MI)
Indiana State Board of Health
Indiana University
Instituto Scientifico HS. Raffaele (Italy)
International Health Management Associates, Inc. (IL)
Jackson Memorial Hospital (FL)
Jersey Shore Medical Center (NJ)
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)
Kenora-Rainy River Regional Laboratory Program (Ontario, Canada)
Kern Medical Center (CA)
Kimball Medical Center (NJ)
King Fahad National Guard Hospital (Saudi Arabia)
King Faisal Specialist Hospital (Saudi Arabia)
King Khalid National Guard Hospital (Saudi Arabia)
King's Daughters Medical Center (KY)
Klinični Center (Slovenia)
LabCorp (NC)
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)
Lewis-Gale Medical Center (VA)
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)
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 Hospitals of Memphis (TN)
MetroHealth Medical Center (OH)
Michigan Department of Community Health
Monmouth Medical Center (NJ)
Monte Tabor – Centro Italo - Brazileiro de Promocao (Brazil)
Montreal Children's Hospital (Canada)
Montreal General Hospital (Canada)
Morton Plant Mease Health Care (FL)
MRL Pharmaceutical Services, Inc. (VA)
MRL Reference Laboratory (CA)
National Institutes of Health (MD)
National University Hospital (Singapore)
Naval Surface Warfare Center (IN)
Nebraska Health System
New Britain General Hospital (CT)
New England Fertility Institute (CT)
New England Medical Center Hospital (MA)
New York Hospital Medical Center of Queens
New York State Department of Health
NorDx (ME)
North Carolina State Laboratory of Public Health
Northern Indiana Education Foundation
North Kansas City Hospital (MO)
North Mississippi Medical Center
North Shore – Long Island Jewish Health System Laboratories (NY)
Northridge Hospital Medical Center (CA)
Northwestern Memorial Hospital (IL)
Ohio Valley Medical Center (WV)
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)
Pocono Hospital (PA)
Polly Ryon Memorial Hospital (TX)
Presbyterian Hospital of Dallas (TX)
Prodia Clinical Laboratory (Indonesia)
Providence Health System (OR)
Providence Seattle Medical Center (WA)
Queen Elizabeth Hospital (Prince Edward Island, Canada)
Queensland Health Pathology Services (Australia)
Quest Diagnostics Incorporated (CA)
Quintiles Laboratories, Ltd. (GA)
Reading Hospital and Medical Center (PA)
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. Barnabas Medical Center (NJ)
St. Boniface General Hospital (Winnipeg, Canada)
St. Elizabeth Hospital (NJ)
St-Eustache Hospital (Quebec, Canada)
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 (CA)
St. Luke's Regional Medical Center (IA)
St. Mark's Hospital (UT)
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 Cabrini Hospital (Montreal, PQ Canada)
Santa Clara Valley Medical Center (CA)
Seoul Nat'l University Hospital (Korea)
Shanghai Center for the Clinical Laboratory (China)
South Bend Medical Foundation (IN)
Southern California Permanente Medical Group
South Western Area Pathology Service (Australia)
Southwest Texas Methodist Hospital
Speciality Laboratories, Inc. (CA)
Stanford Hospital and Clinics (CA)
State of Washington Department of Health
Stormont-Vail Regional Medical Center (KS)
Sun Health-Boswell Hospital (AZ)
Sunrise Hospital and Medical Center (NV)
T.A. Sourasky Medical Center (Israel)
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 Florida
University of Illinois at Chicago
University of the Ryukyus (Japan)
University of Texas M.D. Anderson Cancer Center
University of Virginia Medical Center
University of Washington
UPMC Bedford Memorial (PA)
UZ-KUL Medical Center (Belgium)
VA (Dayton) Medical Center (OH)
VA (Denver) Medical Center (CO)
VA (San Diego) Medical Center (CA)
VA (Tuskegee) Medical Center (AL)
VA Outpatient Clinic (OH)
Vejle Hospital (Denmark)
Viridae Clinical Sciences, Inc. (Vancouver, BC, Canada)
Washoe Medical Center Laboratory (NV)
Watson Clinic (FL)
Wilford Hall Medical Center (TX)
William Beaumont Hospital (MI)
Williamsburg Community Hospital (VA)
Winn Army Community Hospital (GA)
Wishard Memorial Hospital (IN)
Yonsei University College of Medicine (Korea)
York Hospital (PA)
Zale Lipshy University Hospital (TX)

Contents

Contents (Continued)

Foreword

Before a clinical laboratory method can be used for testing patient samples, its analytical performance must be evaluated to demonstrate that the method provides the medically required precision and accuracy. The scope of method evaluation varies according to what organization is performing the evaluation and what is already known about the analytical performance of the method. In order of decreasing amounts of effort, the scopes of evaluation are:

- Evaluation — a measurement of the analytical performance characteristics of a new method by means of laboratory experiments.
- Validation — a) action (or process) of proving that a procedure, process, system, equipment, or method used works as expected and achieves the intended result; or b) confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use can be consistently fulfilled.
- Verification — confirmation by examination of objective evidence that specified requirements have been fulfilled.
- Demonstration — assessing whether a laboratory can follow the manufacturer's instructions and obtain expected results.

The focus of this guideline is a demonstration of analytical precision and accuracy of an analytical method by a laboratory. This guideline is intended to be a companion document to NCCLS document EP5 — *Evaluation of Precision Performance of Clinical Chemistry Devices* and NCCLS document EP9 — *Method Comparison and Bias Estimation Using Patient Samples.* EP5 and EP9 are intended for validating and verifying performance claims. EP15 is intended for demonstrating (rather than "proving") that a laboratory's performance is consistent with these claims.

This guideline has been developed to guide the user through minimum studies necessary to demonstrate that the user can obtain precision and accuracy performance consistent with the manufacturer's claims and, if proficiency testing materials are used, consistent with the test system's peer group as well. It is assumed that the method's performance has previously been evaluated by the manufacturer, using protocols designed to validate and verify performance. It is also assumed that the method being evaluated has been thoroughly evaluated previously in other settings, and that the method and user are inherently capable of the performance claimed by the manufacturer. The experimental and statistical protocols have relatively weak power to reject claims with statistical confidence and therefore should only be used to demonstrate that the method, as performed by the laboratory, is operating in accordance with the manufacturer's claims. This guideline is very limited in scope and is appropriate only for demonstration studies.

The subcommittee had two principal goals during the development of EP15. One was to develop a testing protocol that is simple enough to be applicable in laboratories with a wide variety of sophistication and resources, from the point-of-care or physician's office laboratory to the large clinical laboratory. The second was to develop a protocol that is sufficiently rigorous to provide statistically valid conclusions for demonstration studies. To meet these two needs, the subcommittee developed three- and five-day testing protocols and simplified worksheets for all data gathering, statistical calculations, and tests of observed precision and accuracy.

This document is primarily intended for use when an established method is being initially set up in the laboratory. It provides protocols for demonstrating precision and accuracy. Protocols for validating the manufacturer's suggested reference ranges are included in the most current edition of NCCLS document C28 — *How to Define and Determine Reference Intervals in the Clinical Laboratory*.

There is another NCCLS guideline requiring minimal effort to assess analytical performance: EP10 — *Preliminary Evaluation of Quantitative Clinical Laboratory Methods*. While fairly complex, since it is

based on a multifactor design, it is fairly limited in the amount of data generated. As its title states, EP10 is only appropriate as a preliminary study. EP10 is intended for use in rapid preliminary evaluations of precision, bias, sample carryover, drift, and nonlinearity.

Key Words

Accuracy, demonstration of performance, precision

A Note on Terminology

NCCLS recognizes that harmonization of terms facilitates the global application of standards, and as a matter of organizational policy, is firmly committed to employing terms that are generally used internationally. This initiative includes a mechanism to resolve ISO/CEN/NCCLS differences in nomenclature.

However, NCCLS is also aware 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. Therefore, implementation of this policy must be an evolutionary and educational process that begins with new projects and revisions of existing documents.

Of particular note in EP15-A are several terms whereby NCCLS intends to eliminate confusion, over time, through its commitment to harmonization. These terms and their ISO counterparts include: *Accuracy* vs. *Trueness*; *Analyte* vs. *Quantity; Analytical method* vs. *Measurement procedure; Total precision* vs. *Reproducibility; Within-run precision* vs. *Repeatability; Reportable range* vs. *Measuring range; and Total error* vs. *Error of measurement.* The users of EP15-A should understand that the fundamental meanings of the terms are identical, and to facilitate understanding, the terms are defined along with their ISO counterpart in the guideline's Definitions section.

All terms and definitions will be reviewed for consistency with international use, and revised appropriately during the next scheduled revision of this document.

User Demonstration of Performance for Precision and Accuracy; Approved Guideline

1 Introduction

This guideline has been written to assist the laboratory in bringing an established method (or device or analytical system) on line. It presumes that the method has been checked by the manufacturer and is known to be functioning properly. This guideline provides a minimum implementation protocol necessary to demonstrate that a particular example of a method is operating in accordance with the manufacturer's claims. The laboratory must test the method against these targets for the protocols in this guideline to be applicable.

The guideline is also intended to provide a procedure with which a laboratory can demonstrate acceptable performance as a follow-up to corrective actions taken after a failed proficiency-testing event.

The specific characteristics addressed in this document are precision (within-run and total) and accuracy relative to an accepted standard. Upon successful completion of the protocols recommended in this guideline, the laboratory will have demonstrated that the system is operating in accordance with the manufacturer's claims for precision and accuracy.

This document leads the user through the process of determining the match between the laboratory's actual performance and the expected performance of the method. If the laboratory's performance is not consistent with the expected level of performance, remedial actions will probably be required.

Underlying this protocol is an assumption that the user can operate the method properly and obtain the performance claimed by the manufacturer.

1.1 Scope

Prior to selecting a method for an analyte and evaluating that method's analytical performance, the laboratory must establish minimum performance specifications for the method based on the laboratory's clinical and proficiency testing needs. Lists of medically based analytical performance standards are given in the references.[1-4] Some regulatory and accrediting programs[a] (e.g., CLIA, CAP) specify minimum performance standards, most frequently for proficiency testing. If regulatory performance standards apply, these define the minimum performance the method must achieve. These standards are expressed in terms of total allowable difference (total error) from a reported group mean value or true value of the concentration of the analyte. Precision and accuracy goals in terms of allowable standard deviation and bias must be derived from allowable total error. Discussions of the relationship between allowable error and allowable standard deviation and bias are included in some of the publications listed in the references.[1-4]

For the performance characteristics evaluated in this document, the following performance goal formats are recommended in order to conform to the evaluation result formats:

Precision. Precision goals should be stated as the maximum allowable SD and/or CV at each analyte concentration to be tested.

[a] For example, in the U.S., CLIA and CAP.

Accuracy. Accuracy goals should be stated as maximum allowable bias at each analyte concentration to be tested. Maximum allowable bias may be expressed as either an absolute concentration or as a percentage of the concentration.

Total Error. Total error goals should be stated as the maximum permissible difference between an individual specimen's result and the target value for that specimen. The target value may be determined by:

(a) the method's peer group in proficiency testing

(b) an assigned reference method in proficiency testing

(c) a comparative method in a comparison of patient samples experiment

(d) the manufacturer of a reference material

Ideally the laboratory can select a method whose manufacturer's claims for precision and accuracy are within the limits of the performance standards specified by the laboratory. Other factors in this selection include application characteristics such as cost of operation, sample size, turnaround time, etc.

The approach taken in this protocol is to demonstrate that observed performance is consistent with the manufacturer's claims. When multiple analytes are tested on a single analytical system, it may not be possible to select a single system, which has acceptable claimed performance for every analyte. Although the performance characteristics of most products fall within manufacturers' claims, when a product's claimed performance does not meet performance goals, this protocol is not appropriate for demonstrating performance. Other, more rigorous NCCLS protocols should be employed to validate the methods performance against the user's needs.

This document provides experimental protocols and data analysis procedures designed to enable a user laboratory to demonstrate that it has obtained analytical performance comparable to that established by the manufacturer of an *in vitro* diagnostic device. Generally this guideline will be employed when a new device is being evaluated prior to its application for routine testing in the user's laboratory.

As this guideline is very limited in scope, it is not intended for validation or verification of the analytical performance of a diagnostic device. Other NCCLS guidelines have been developed for that purpose. This guideline has been developed for use in situations in which the performance of the device has been previously established and documented by experimental protocols of much larger scope and duration. Most often such documentation has been provided by the manufacturer and is part of the product labeling. The present guideline is to be employed by the user laboratory to demonstrate that it has obtained performance consistent with that documented by the manufacturer.

Accreditation and regulatory agencies require laboratories to establish performance specifications for each analytical method, and to verify or demonstrate that the method's analytic performance meets these specifications. This guideline specifically deals with demonstrating analytic precision and accuracy.

More complete evaluations of precision and accuracy (as would be required to validate or verify the performance of a newly developed method), as well as guidelines for developing or validating reference ranges and for evaluating sensitivity and specificity are given in related NCCLS publications listed at the end of this document.

1.2 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 the most current edition of NCCLS document M29—*Protection of Laboratory Workers from Occupationally Acquired Infections.*

2 Overview of the Protocol

All of the experimental work in this protocol can be completed in five days or less.

2.1 Device Familiarization Period (Section 4)

The device familiarization period is the time given to operators to become both familiar and comfortable with the details of the instrument's operation and the assay procedure. Including a familiarization period into the time line for an evaluation study is critical for meaningful evaluations of precision. If the operator has not had the opportunity for a familiarization period, including the opportunity to perform the assay prior to beginning the precision protocol, the first data points generated by the operator may cause the laboratory to assume the test system has a higher level of imprecision than is actually the case.

The familiarization period is also the time to verify that the QC materials the laboratory intends to use for the assay perform as expected.

2.2 Precision Evaluation Experiment (Section 5)

The precision evaluation experiment provides the user with guideline procedures for demonstrating precision performance. Usually the manufacturer makes two types of precision claims—within-run precision (σ_{within}) and total within-laboratory precision (σ_{total}). This section provides statistical methods for identifying gross deviations from both types of claims. Two protocols are provided, based on a statistical power evaluation and based on the information provided by the manufacturer about the difference between σ_{within} and σ_{total}.

2.3 Accuracy Evaluation Experiment (Section 6)

The accuracy evaluation experiment provides the user with two different approaches. Either or both may be used.

(a) Comparability. Accuracy may be assessed by way of a split-sample comparison experiment, by analyzing 20 patient samples well distributed over the entire test range. Results from the two methods (the method under evaluation and a comparative method) are compared to determine if significant bias exists.

(b) Recovery of expected values from assayed reference materials. Accuracy may be assessed by analyzing proficiency test materials and other assayed materials, and comparing the results for the method under evaluation to those obtained by the method's peer group.

2.4 Reportable Range and Reference Ranges

User demonstration of reportable range is included in the most current version of NCCLS document EP6 — *Evaluation of the Linearity of Quantitative Analytical Methods*. User verification of reference ranges is included in the most current version of NCCLS document C28 — *How to Define and Determine Reference Intervals in the Clinical Laboratory*. These topics are not covered in this document.

3 Definitions[b]

For the performance characteristics evaluated in EP15-A, the following terms and their definitions are appropriate. Please note that the terms appear along with their ISO counterpart and corresponding definition, where appropriate.

Accuracy, *n* - The term **Accuracy** is defined the way ISO[5] defines the term **Trueness,** i.e., the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value.

Analyte, *n* **-** The term **Analyte** is defined the way VIM[6] defines the term **Quantity**, i.e., Attribute of a phenomenon, body, or substance that may be distinguished qualitatively and determined quantitatively.

Analytical method, *n* **-** This term is defined the way VIM defines **Measurement procedure;** i.e., set of operations, described specifically, used in the performance of particular measurements according to a given method.

Peer group, *n* - In proficiency testing, a group of presumably identical test systems.

Reportable range, *n* - The range of values (in units appropriate for the analyte) over which the acceptability criteria for the method have been met; that is, where errors due to nonlinearity, imprecision, or other sources are within defined limits; **NOTE:** This is similar to the VIM definition for Measuring range or working range, i.e., a set of values of measurands for which the error of a measuring instrument is intended to lie within specified limits.

Total error, *n* **-** The difference between an individual specimen's result and the target value for that specimen; **NOTE:** This is similar to the VIM definition for **Error of measurement,** i.e., the result of a measurement minus a true value of the measurand.

Total precision, *n* - **Total precision** is defined the way VIM defines **Reproducibility** (of results of measurements); i.e., closeness of the agreement between results of measurements of the same measurand carried out under changed conditions of measurement.

Within-run precision, *n* - **Within-run precision** is defined the way VIM defines **Repeatability** (of results of measurements); i.e., closeness of the agreement between results of successive measurements of the same measurand carried out under the same conditions of measurement.

[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.

4 Familiarization Period

After the system has been checked out by the manufacturer, staff must become familiar with the operation, maintenance procedures, methods of sample preparation, calibration, and monitoring functions. The length of time required for this process is variable, depending on the complexity of the device. Calibration should be verified during this period, if appropriate. At the end of this time, the operator(s) should be confident in the operation of the device.

4.1 Operator Training

The operation, maintenance procedures, methods of sample preparation, and calibration and monitoring functions must be learned. Some manufacturers provide this training. The device should be set up and operated in the individual laboratory long enough to understand all of the procedures involved to avoid problems during the evaluation of its performance. Training should include the use of actual sample material (including pools, controls, leftover patient specimens, or any other test materials appropriate for the device.

All possible contingencies (such as error flags, error correction, calibration, etc.) that might arise during routine operation should be carefully monitored. Data should not be collected during this period. Operator training is not complete until the user can confidently operate the device.

4.2 Quality Control Procedures

Quality control procedures to be followed during the evaluation are established during the familiarization period. It is important to establish that the device is operating in accordance with the manufacturer's specifications. To demonstrate this during an evaluation, use the control procedures recommended by the manufacturer. Due to the short duration of this protocol, the estimated standard deviations should not be used by themselves to establish quality control ranges. For guidance on establishing ongoing quality control procedures, refer to the most current edition of NCCLS document C24—*Statistical Quality Control for Quantitative Measurements: Principles and Definitions.*

4.3 Materials for Precision Experiments

Materials to be used as samples for precision experiments are tested during this period to verify that they perform as expected. Since precision is directly related to concentration, analyte concentrations should be focused at or near medical decision points. For example, a glucose method's performance should be assessed at a concentration near 126 mg/dL, above which concentration, a fasting glucose result may indicate disease. For certain analytes, it may also be important to measure precision at the upper and lower limits of the reportable range. Normally, it is sufficient to select materials that have analyte values near the concentrations the manufacturer used to establish the precision claims for the assay. (This information is in the package insert or instructions for use provided by the manufacturer of the test system under evaluation.) Acceptable materials for precision experiments include control samples (other than those used to assess whether the assay is in control), standards, previously assayed patient samples, or suitable materials that have a known value. The materials used for precision experiments should mimic the matrix of the patient sample. For example, an assay for a whole blood analyte should use a material for precision experiments that is as close as possible to human whole blood.

5 Demonstration of Precision Performance

Precision is a quantitative value indicating the extent of dispersion of a set of replicate measurements. Precision can be reported either as a standard deviation (SD) or a coefficient of variation (CV), which expresses the standard deviation as a percentage of the mean value of the replicate measurements.

In either case, the mean value should be reported also. Increasing values of the standard deviation or coefficient of variation indicate decreasing precision of the measurements.

Precision is generally considered as either within-run precision or total precision. Within-run precision is a quantitative value indicating the extent of dispersion of a set of replicate measurements of an analyte when all measurements are made within a single run of a method. A run is the period of time over which the performance of the method is expected to be stable.

Total precision is a quantitative value indicating the extent of dispersion of replicate measurements of an analyte over a longer time, when all known, major sources of measurement error are accounted for. Total precision is measured as the sum of various measurement error sources, including the within-run precision.

The ratio of within-run precision to total precision affects the experimental design needed to demonstrate that a laboratory's total precision is consistent with a manufacturer's imprecision claims.

5.1 Experimental Design — Numbers of Days and Replicates

NOTE: The lower case sigma (σ), combined with a subscript of "within" or "total" will be used to designate the manufacturer's claimed values of within-run standard deviation and total standard deviation, respectively. The lower case "s," combined with a subscript of "within" or "total" will be used to designate the user's estimated values of within-run standard deviation and total standard deviation, respectively.

The experimental protocols described below are appropriate for two analyte levels and a five percent statistical false rejection rate. (Please see Appendix C for detailed information on statistical false rejection rates.) Formulas to determine the appropriate experimental designs to verify imprecision claims for more than two analyte levels or other error rates are given in Appendix C. The appropriate experimental protocol depends on the relative size of σ_{within} to σ_{total}. Assuming a five percent statistical false rejection rate:

(a) If $\sigma_{within} < 2/3\ \sigma_{total}$, the experiment should last for five days and consist of four replicates per level per run. THIS PROTOCOL IS APPROPRIATE WHEN THE BETWEEN-DAY VARIATION IS AN IMPORTANT COMPONENT OF THE TOTAL IMPRECISION.

(b) If $\sigma_{within} > 2/3\ \sigma_{total}$, the experiment should last for three days and consist of three replicates per level per run. THIS PROTOCOL IS APPROPRIATE WHEN THE BETWEEN-DAY VARIATION IS NOT AN IMPORTANT COMPONENT OF THE TOTAL IMPRECISION.

(c) If the relationship of σ_{within} to σ_{total} is not known, the five-day experiment should be performed, with four replicates per level.

5.2 Specific Procedures

(1) Analyze one run per day:

- For $\sigma_{within} < 2/3\ \sigma_{total}$, analyze four replicate samples at each of two concentrations daily for five days.
- For $\sigma_{within} > 2/3\ \sigma_{total}$, analyze three replicate samples at each of two concentrations daily for three days.

(2) If a run must be rejected because of quality control procedures or operating difficulties, discard the data, and conduct an additional run.

(3) Include the daily quality control samples normally used (See Section 4.2).

(4) Samples for the accuracy experiment may be tested in the same runs.

(5) Calibrate as specified in the manufacturer's instructions for operators. If the manufacturer indicates in its claim that its precision data was generated over multiple calibration cycles, then the operator may choose to recalibrate during the experiment.

5.3 Recording the Data

Appendix A contains examples of data recording sheets to summarize data. This type of summary is valuable in the statistical analysis described below. Blank worksheets are included for the five-day and three-day protocols. Examples of completed worksheets are included in Appendix B.

5.4 Calculation of Precision Estimates

After collecting the data and transcribing them onto an appropriate recording sheet, the calculations described in this section should be performed. Blank worksheets for the calculations are included in Appendix A. Examples of completed calculation worksheets for both protocols are included in Appendix B.

Separate calculations should be performed for each level of concentration.

5.4.1 Within-Run Precision

Calculate the within-run precision from the following formula (see worksheet in Appendix A).

$$s_{within} = \sqrt{\frac{\sum_{d=1}^{D}\sum_{i=1}^{n}(x_{di} - \bar{x}_d)^2}{D(n-1)}}$$

where:

Σ indicates that the terms to the right of Σ are to be summed (see worksheet in Appendix A),
D = total number of days (3 or 5),
n = total number of replicates per day (3 or 4),
x_{di} = result for replicates per day (3 or 4 replicates), and
$\bar{x}_d$ = average of all results for day d.

It should be noted that if the experiment protocol is not followed exactly (same number of replicates for all runs on separate days), the within-run precision estimate will be incorrect.

5.4.2 Total Precision

Total precision is calculated using the following formulas (see worksheet in Appendix A). These calculations are based upon the variance components method discussed in NCCLS document EP5—*Evaluation of Precision Performance of Clinical Chemistry Devices*.

Calculate the variance term, B, for the daily means from the formula (see worksheet in Appendix A):

$$B = \frac{\sum_{d=1}^{D} (\bar{x}_d - \bar{\bar{x}})^2}{D-1}$$

where:

$\bar{x}_d$ = average of all results for day d, and

$\bar{\bar{x}}$ = average of all results.

Calculate s_{total} from the formula (see worksheet in Appendix A):

$$s_{total} = \sqrt{\frac{n-1}{n} \bullet s^2_{within} + B}$$

where n = number of replicates per run (either three or four depending on protocol).

5.5 Comparison of Estimated Within-Run Precision to Within-Run Precision Claims

Verify within-run precision claims by comparing the within-run precision estimate calculated in Section 5.4.1 to the manufacturer's claim. If the manufacturer's claim for within-run precision is in terms of coefficient of variation, convert the coefficient of variation into a standard deviation at the average concentration of all results for the material tested:

$$\sigma_{within} = CV_{within} \bullet \bar{\bar{x}}$$

where CV_{within} is the manufacturer's claimed within-run coefficient of variation.

If the estimated within-run standard deviation is less than the manufacturer's claimed standard deviation, the user has demonstrated precision consistent with the claim. If the within-run standard deviation is greater than the manufacturer's claimed within-run standard deviation, note that a user's within-run standard deviation can be larger than the manufacturer's claim and not be statistically different from the claim. If the calculated within-run standard deviation is larger than the manufacturer's claimed within-run standard deviation, test whether it is statistically significantly larger (really different) using the following four steps:

(1) Calculate the within-run precision degrees of freedom, ν. For an experiment with D days duration and n replicates per run, ν is equal to D • (n-1). For the recommended protocols of three and five days duration:

ν= 6 and 15, respectively.

(2) Determine the $(1-\alpha/\ell)$ percent point, C, of the χ^2 ("Chi-Squared") distribution with ν degrees of freedom. Here, α is the false rejection rate (usually 5%), and ℓ is the number of levels tested. The percent points for C corresponding to two, three, and four levels of testing are 97.5%, 98.33%, and 98.75%, respectively. Table 1 lists the values of C for these percentage points; other values can be obtained from any standard statistics book or most commonly used computer spreadsheet programs. For the recommended protocols of three and five days duration, with two levels, C = 14.45 and 27.49, respectively.

(3) Calculate the verification value as

$$\frac{\sigma_{within} \bullet \sqrt{C}}{\sqrt{v}}$$

(4) If the estimated within-run precision, s_{within}, is less than the verification value, the manufacturer's claim for within-run precision is verified. If the claim is not verified, then the user should contact the manufacturer for diagnostic assistance.

Table 1. Selected Percentage Points of the Chi-Square Distribution for Selected Numbers of Levels to Provide 5% False Rejection Rate

	Number of Levels		
Degree of Freedom	2	3	4
3	9.35	10.24	10.86
4	11.14	12.09	12.76
5	12.83	13.84	14.54
6	14.45	15.51	16.24
7	16.01	17.12	17.88
8	17.53	18.68	19.48
9	19.02	20.21	21.03
10	20.48	21.71	22.56
11	21.92	23.18	24.06
12	23.34	24.63	25.53
13	24.74	26.06	26.98
14	26.12	27.48	28.42
15	27.49	28.88	29.84
16	28.85	30.27	31.25
17	30.19	31.64	32.64
18	31.53	33.01	34.03
19	32.85	34.36	35.40
20	34.17	35.70	36.76
21	35.48	37.04	38.11
22	36.78	38.37	39.46
23	38.08	39.68	40.79
24	39.36	41.00	42.12
25	40.65	42.30	43.35

5.6 Comparison of Estimated Total Precision to Total Precision Claims

Verify total precision claims by comparing the total-run precision estimate calculated in Section 5.4.2 to the manufacturer's claim. If the manufacturer's claim for total precision is in terms of coefficient of variation, convert the coefficient of variation into a standard deviation at the average concentration of all results for a the material tested:

$$\sigma_{total} = CV_{total} \bullet \bar{\bar{x}}$$

where CV_{total} is the manufacturer's claimed total coefficient of variation.

If the estimated total standard deviation is less than the manufacturer's claimed total standard deviation, the user has demonstrated precision consistent with the claim. Go to Section 6 — "Demonstration of Accuracy." If the total standard deviation is greater than the manufacturer's claimed standard deviation, note that a user's total precision can be larger than the manufacturer's claim and not be statistically different from the claim. If the calculated total standard deviation is larger than the manufacturer's claimed total standard deviation, test whether it is statistically significantly larger (really different) using the following four steps:

(1) Calculate the total precision degrees of freedom, T. For an experiment with D days duration and n replicates per run (see worksheet in Appendix A):

$$T = \frac{((n-1)s^2_{within} + (nB))^2}{(\frac{n-1}{D})s^4_{within} + (\frac{n^2B^2}{D-1})}$$

where B is calculated in Section 5.4.2.

(2) Determine the $(1-\alpha/\ell)$ percent point, C, of the χ^2 ("Chi-Squared") distribution with ν degrees of freedom. Here, α is the false rejection rate (usually 5%), and ℓ is the number of levels tested. The percent points for C corresponding to two, three, and four levels of testing are 97.5%, 98.33%, and 98.75%, respectively. Table 1 lists the values of C for these percentage points; other values can be obtained from any standard statistics book or most commonly used computer spreadsheet programs.

(3) Calculate the verification value as

$$\frac{\sigma_{total}}{\sqrt{T}}\sqrt{C}$$

(4) If the total precision estimate, s_{total}, is less than the verification value, the user has demonstrated total precision consistent with the manufacturer's claim. If s_{total} exceeds the verification value, the user has not demonstrated total precision consistent with the manufacturer's claim, and the user should contact the manufacturer for diagnostic assistance.

6 Demonstration of Accuracy

Accuracy is conformance to a true measure, accepted standard, or expected value. For a test result, accuracy is the difference between the test result and the true value for an analyte. For an analytical process, accuracy is expressed as the difference between the average result obtained by a method under specified conditions and the result accepted as true or standard.

This protocol has provisions for demonstrating accuracy by two procedures (see Appendixes D and E):

(1) Comparison of patient specimens results to another method. When possible, the comparison of patient specimens experiment should be performed. This technique avoids various artifacts, which can be present with commercial reference materials, and some methods. Comparison of patient specimens is particularly important for initial evaluation of a method in a laboratory. It is strongly recommended when close agreement between methods is expected. The comparison of patient specimens can also form the basis of establishing the relationship between multiple methods. This will ensure the laboratory's ability to provide equivalent results irrespective of which method is used to assay a patient specimen (see NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples*).

(2) Recovery of expected values for assayed reference materials. Once experience is gained with a method, the reference material strategy may be more convenient for validation following a calibration or proficiency test failure.

6.1 Comparison of Patient-Specimen Results to Another Method

The choice of the comparative (reference) method is critical for the interpretation of the results of this experiment.

If the user intends to demonstrate accuracy consistent with a manufacturer's claim, the comparative method must be the same method used by the manufacturer as the comparative method in developing the accuracy claim. Often, when the new method is a revision of a previous method of the same manufacturer, or application of the manufacturer's previous method to a new instrument, the user's current method is the manufacturer's comparative method. The manufacturer's accuracy claim is applicable in this instance and should be used as the basis for demonstrating accuracy in this experiment.

Often, the current method is different from the comparative method, or is a reference method or laboratory, and the user intends to demonstrate accuracy of the new method relative to a method different from the one used by the manufacturer as the comparative method. In this case, the manufacturer's claim may not be valid, and should not be the basis for demonstrating accuracy. The user must specify a medically allowable bias between the new method and the comparative method, and use this as the basis for demonstrating accuracy. Guidance for specifying allowable bias can be found.[1-3] The experimental protocol for comparing results of patient specimen testing in this guideline has been designed as a demonstration protocol. To keep the experimental work simple, each patient is tested singly by the comparative test methods. Because it has relatively low power to detect bias between methods, it should only be employed when the user anticipates close agreement between the methods. Otherwise, the user should employ NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples* because it includes more patient specimens and testing in duplicate by both methods.

Once a comparison method has been selected, the following steps should be followed. See worksheet in Appendix D and example completed worksheet in Appendix E.

(1) Obtain 20 patient specimens whose analyte concentrations span the reportable range of the test method. Do not use specimens whose analyte concentrations exceed the reportable range. The sample type must be compatible between the comparative and test methods. Exclude specimens containing substances identified as interferents (in the manufacturer's instructions) for either the test method or comparative method. Specimens should be assayed as close to the day of collection as possible to mimic conditions expected when the test method is in regular use. Follow the manufacturer's instructions for collection and handling of patients' specimens to be assayed on the test system. For laboratories in which abnormal specimens are infrequently observed, it may be necessary to store them until there is a sufficient number for the experiment. If stored specimens are used, they should be refrigerated, if consistent with analyte stability and manufacturer's instructions,

to avoid possible artifacts introduced in the freeze-thaw cycle. Some laboratories may need to use frozen samples, particularly if the duration of the experiment and the manufacturer's product labeling require that frozen samples must be used, and they must be well mixed and examined for particulate matter after thawing and before use. If particulate matter is detected, they should be centrifuged, and the supernatant should be used for sampling. If stored specimens must be used, testing by the comparative and test methods should be performed within an hour or two if possible. The specimens' test results should be evenly spread over the measurement range. If the full range cannot be accommodated, the conclusions will only be applicable to the range tested. A separate validation of measurement range may allow adequate conclusions regarding accuracy based on a restricted range of specimen results.

(2) Assay the specimens on both the test and comparative methods. Assays by each method should be completed within four hours of each other on the same day. Conclusions will be most reliable if the assays are performed on five to seven specimens per day for three to four days. This testing may be combined with precision testing. Performing assays on several days allows averaging of any between-day variability, which may exist for either method. Examine the results after each assay event. If an isolated specimen's results for the test and comparative methods differ more than observed for other specimens, retest that specimen in duplicate on both methods. Some patient specimens may show unexpectedly large differences between methods due to differences in specificity between methods. If the difference persists after repeat testing in duplicate on both methods, it represents a real analytical difference, which exists between the two methods. The manufacturer may be able to provide additional clarification regarding method specificity. The data are kept for subsequent analysis. If the difference is not confirmed by repeat testing, use the first pair of repeat values for data analysis.

(3) Appropriate quality control procedures should be followed for each method. Any unacceptable analytical performance should be corrected, and the specimens from that run should be retested.

(4) Calculate the difference (or individual specimen bias) in reportable units and/or the percent difference (or percent individual specimen percent bias) between each specimen's results for the two methods.

Individual specimen bias in reportable units = b_i = (test result$_i$ - comparison result$_i$)

Individual specimen bias in percent = $\%b_i = 100 \bullet$ ([test result$_i$ - comparison result$_i$]/comparison result$_i$)

Construct a plot of bias and/or percent bias (vertical axis) vs. comparison method result (horizontal axis) for each specimen. Examine the bias plot to determine if the difference between methods is relatively constant over the concentration range tested. If constant bias vs. concentration or constant percent bias vs. concentration is observed, then the mean bias in step (5) represents the average difference between the methods. This value is compared to the manufacturer's claim to demonstrate method accuracy.

If neither the bias nor the percent bias is constant over the concentration range tested, the data should be partitioned into two segments and the average bias calculated separately for each segment. If the bias shows a progressively changing relationship to concentration, no average bias can be calculated. In this case, more data will be needed to validate the accuracy of the method. Refer to NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples* for additional information.

(5) Calculate the bias in reportable units and/or percent between the two methods.

$$\bar{b} = \frac{\sum_{i=1}^{I} b_i}{n}$$

$$\overline{\%b} = \frac{\sum_{i=1}^{I} \%b_i}{n}$$

(6) Calculate the standard deviations of the bias and/or bias in percent.

$$s_{\bar{b}} = \sqrt{\frac{\sum_{i=1}^{I} (b_i - \bar{b})^2}{n-1}}$$

$$s_{\overline{\%b}} = \sqrt{\frac{\sum_{i=1}^{I} (\%b_i - \overline{\%b})^2}{n-1}}$$

6.1.1 Procedure for Demonstration of Accuracy by Comparison of Patient Specimens

If the bias or percent bias is less than the manufacturer's claimed bias or percent bias, the user has demonstrated bias consistent with the claim. Go to Section 6.2 — "Recovery of Expected Values from Assayed Reference Materials" to further demonstrate accuracy of recovery, if desired. If the bias or percent bias is greater than the manufacturer's claimed bias or percent bias, note that a user's bias or percent bias can be larger than the manufacturer's claim and not be statistically different from the claim. If the bias or percent bias is larger than the manufacturer's claimed bias, test whether it is statistically significantly larger (really different) using the following four steps.

NOTE: This method can be used to test the bias only when the bias or percent bias is constant throughout the reportable range, or to test the biases or percent biases calculated by splitting the data as described in Section 6.1 (4).

Verification of the manufacturer's claim is performed using the following steps:

(1) Assume a false rejection rate, α. Typical values selected for this error rate are 1% and 5%.

(2) Determine the (100 - α) percent point, *t,* of the t-distribution with n-1 degrees of freedom. Here n represents the number of patient samples. For example, if α equals 1% and n equals 20, the (100 - α) point of the t-distribution with 19 degrees of freedom is 2.539. Other values of *t* can obtained from any standard statistics book[7] or most commonly used computer spreadsheet programs for different values of α and n.

(3) Calculate the verification value for bias in reportable units as

$$\frac{t \bullet s_{\bar{b}}}{\sqrt{n}} + \beta$$

where β is the manufacturer's claimed value of bias.

If the estimated bias $\bar{b}$ is less than the verification value, the user has demonstrated bias consistent with the manufacturer's claim.

If percent bias is used, calculate the verification value for percent bias as

$$\frac{t \bullet s_{\overline{\%b}}}{\sqrt{n}} + \beta$$

where β is the manufacturer's claimed value of percent bias.

If the estimated percent bias $\overline{\%b}$ is less than the verification value, the user has demonstrated percent bias consistent with the manufacturer's claim.

6.2 Recovery of Expected Values from Assayed Reference Materials

Reference materials with analyte target values are available from several vendors. These materials are typically manufactured from human source materials but contain additives to achieve desired analyte levels and are processed to promote stability. Because of manufacturing requirements, these processed materials have a solution matrix different from that of an authentic human specimen. The difference in matrix may cause an altered analytical response, which is unique to a particular material–method combination. Consequently, it is incorrect to use an analyte target value assigned by a definitive or reference method unless the reference material has been specifically evaluated and found suitable for use with a method. Reference materials, which have target values assigned specifically for the method of interest, can be used for demonstration of accuracy. In this case the accuracy demonstration is limited to confirming that the test method performs similarly to that same method as used in other laboratories. Another consequence of material–method matrix interactions is the potential for different response with a lot of reagent not represented in the group of laboratories used for assignment of the target value. If a reagent lot difference occurs, the method-specific target value may not be appropriate for the new lot of reagents. The method manufacturer should be consulted for advice regarding appropriate reference materials for validation of accuracy.

Some sources of value-assigned materials, which can be used for accuracy validation, are listed below:

(a) Fresh frozen human serum or other nonadulterated human materials. Such materials are available from NIST and CAP with NRSCL reference and definitive values assigned for some analytes.

(b) Reference materials derived from proficiency testing programs. These materials are value assigned by a large number of laboratories and frequently represent numerous lots of reagents and system calibrators. Consequently, their target values represent average performance for the method.

(c) Materials provided by the method manufacturer for accuracy validation or quality control. These materials have been specifically designed for the analytical system being tested, but are generally not suitable for use on another manufacturer's method.

(d) Materials used in interlaboratory quality control programs are assayed by a relatively large number of laboratories, and their peer group mean values can be used to assess agreement. Caution should be

exercised that an adequate number of laboratories is included in the peer group for a reliable mean value. Ten participants in a method group is generally considered the minimum for a reliable mean value. These programs may sample a small number of reagent lots for a method, which can affect the reliability of the target values for a new lot of reagents in an individual laboratory.

(e) Materials provided by third-party vendors, which have been value-assigned for a number of different methods. These are similar to proficiency testing or regional quality control materials but generally have far fewer laboratories contributing to the peer group mean value. Consequently the target value is less reliable. In addition, relatively few different lots of reagents may have been sampled, which can further affect the reliability of the assigned target values.

6.2.1 Procedure for Demonstration of Accuracy with Reference Materials

(1) Select the best available materials suitable for the method. A minimum of two analyte levels should be tested, although five are preferred to simulate a proficiency testing event, and more may be preferable to adequately evaluate the full measurement range. The levels selected should represent the low and high measurement range for the method. It may also be useful to test intermediate values corresponding to important medical decision levels. Use caution that the levels selected represent analyte values, which can be measured with good precision by the test method.

(2) Prepare materials according to the manufacturer's instructions. Take precautions that the materials are thoroughly mixed prior to use.

(3) Assay each material. Although one replicate is acceptable, duplicate assays will provide more information at minimal cost.

6.2.2 Acceptance Test for Demonstration of Accuracy with Reference Materials

(1) If proficiency testing is available for this analyte, interpret results of this experiment using the acceptance criteria of the proficiency test provider. If reference materials were analyzed in replicate, compare individual values to limits for acceptable performance.

(2) If no proficiency testing is available for this analyte, compare the differences between the laboratory result and the target value for each material to the specification for medically allowable total error.

6.2.3 Procedure for Demonstration of Accuracy with Manufacturer-Provided Reference Materials

Select and prepare materials as directed by the manufacturer. Assay each material in replicate, as directed by the manufacturer.

6.2.4 Acceptance Test for Demonstration of Accuracy with Manufacturer-Provided Reference Materials

Follow the manufacturer's instructions for calculations and interpretation of the data.

References

[1] Garber CC, Carey RN. Evaluation of methods. In: Kaplan LA, Pesce AJ, eds. *Clinical Chemistry: Theory, Analysis, and Correlation*. 3rd ed., St. Louis: Mosby; 1996:402-423.

[2] Koch DD, Peters Jr. T. Selection and evaluation of methods. In: Burtis CA, Ashwood ER, eds. *Tietz Textbook of Clinical Chemistry*. 3rd ed. Philadelphia: W.B. Saunders Co; 1999:508-526.

[3] Fraser CG, Hyltoft Petersen P, Ricos C, Haeckel R. Quality specifications. In: Haeckel R, ed. *Evaluation Methods in Laboratory Medicine*. Weinheim: VCH; 1993:86-100.

[4] Westgard JO, Burnett RW. Precision requirements for cost-effective operation of analytical processes. *Clin Chem*. 1990;36:1629-1632.

[5] International Organization for Standardization (ISO). *Statistics—Vocabulary and Symbols*. ISO 3534-1. Geneva:1993.

[6] International Organization for Standardization (ISO). *International Vocabulary of Basic and General Terms in Metrology (VIM)*. Geneva: 1993.

[7] ASQC. *Glossary & Tables for Statistical Quality Control*. Milwaukee, WI. ASQC Quality Press; 1983:1160.

Appendix A. Sample Data Recording Sheets – Precision Experiment

Use a separate sheet for each concentration.

Data Sheet #1 (for use when $\sigma_{within} \leq 2/3\ \sigma_{total}$)

Device _____ Analyte _____ Concentration _____

Reagent Source/Lot _____ Calibrator Source/Lot _____

	Run 1	Run 2	Run 3	Run 4	Run 5
Date/Operator					
Replicate 1 (x_1)					
Replicate 2 (x_2)					
Replicate 3 (x_3)					
Replicate 4 (x_4)					
$\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$					
$\bar{x}_d = \frac{\sum_{i=1}^{4} x_i}{4}$					
$x_1 - \bar{x}_d$					
$(x_1 - \bar{x}_d)^2$					
$x_2 - \bar{x}_d$					
$(x_2 - \bar{x}_d)^2$					
$x_3 - \bar{x}_d$					
$(x_3 - \bar{x}_d)^2$					
$x_4 - \bar{x}_d$					
$(x_4 - \bar{x}_d)^2$					
$\sum_{1}^{4} (x_i - \bar{x}_d)^2 = (x_1 - \bar{x}_d)^2 + (x_2 - \bar{x}_d)^2 + (x_3 - \bar{x}_d)^2 + (x_4 - \bar{x}_d)^2$					
$sd_{run}^2 = \frac{\sum_{i=1}^{4} (x_i - \bar{x}_d)^2}{3}$					
$\bar{x}_d - \bar{\bar{x}}$ (see below)					
$(\bar{x}_d - \bar{\bar{x}})^2$					

1. Calculation of Grand Mean = $\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4 + \bar{x}_5}{5} =$ _____

2. Calculation of s_{within} :

$$sd^2_{run,average} = \frac{sd^2_{run\ 1} + sd^2_{run\ 2} + sd^2_{run\ 3} + sd^2_{run\ 4} + sd^2_{run\ 5}}{5} =$$ _____

$$s_{within} = \sqrt{sd^2_{run,\ average}} =$$ _____

3. Calculation of s_{total}

s_{within} = _____ , from above

$$B = \frac{\sum_{d=1}^{5} (\bar{x}_d - \bar{\bar{x}})^2}{4} = \frac{(\bar{x}_1 - \bar{\bar{x}})^2 + (\bar{x}_2 - \bar{\bar{x}})^2 + (\bar{x}_3 - \bar{\bar{x}})^2 + (\bar{x}_4 - \bar{\bar{x}})^2 + (\bar{x}_5 - \bar{\bar{x}})^2}{4} =$$ _____

$$s_{total} = \sqrt{\frac{n-1}{n} \bullet s^2_{within} + B}$$, where n = number of replicates in each run.

$$s_{total} = \sqrt{\frac{3}{4} \bullet s^2_{within} + B} =$$ _____

4. Verification of Within-Run Precision Claim:

Compare calculated s_{within} to claimed σ_{within}:

If calculated s_{within} < claimed σ_{within}, within-run precision has been demonstrated to be consistent with the manufacturer's claim. Go to paragraph 5, "Verification of Total Precision Claim."

If calculated s_{within} > claimed σ_{within}, note that a user's within-run precision can be larger than the manufacturer's claim and not be statistically different from the claim.

Calculate the verification value

$\nu = 15$, $C = 27.49$

Verification value = $\frac{\sigma_{within} \bullet \sqrt{C}}{\sqrt{\nu}} = \sigma_{within} \bullet 1.354 =$ _____

Compare calculated s_{within} to verification value:

If calculated s_{within} < verification value, within-run precision has been demonstrated to be consistent with the manufacturer's claim.

If calculated s_{within} > verification value, within-run precision has not been demonstrated to be consistent with the manufacturer's claim; contact the manufacturer for help.

5. Verification of Total Precision Claim:

Compare calculated s_{total} to claimed σ_{total}:

If calculated s_{total} < claimed σ_{total}, total precision has been demonstrated to be consistent with the manufacturer's claim. Go to Section 6, "Demonstration of Accuracy."

If calculated s_{total} > claimed σ_{total}, note that a user's total precision can be larger than the manufacturer's claim and not be statistically different from the claim.

Calculate degrees of freedom, T:

$$T = \frac{((n-1)s^2_{within} + (nB))^2}{(\frac{n-1}{D})s^4_{within} + (\frac{n^2B^2}{D-1})},$$

where D = number of days, and
n = number of replicates.

$$T = \frac{(3s^2_{within} + 4B)^2}{0.6s^4_{within} + 4B^2} = \_\_\_\_\_$$

C = _____ (obtain from Table 1 for T degrees of freedom)

$$\text{Verification value} = \frac{\sigma_{total} \bullet \sqrt{C}}{\sqrt{T}} = \_\_\_\_\_$$

Compare calculated s_{total} to verification value:

If calculated s_{total} < verification value, total precision has been demonstrated to be consistent with the manufacturer's claim. Go to Section 6, "Demonstration of Accuracy."

If calculated s_{total} > verification value, total precision has not been demonstrated to be consistent with the manufacturer's claim; contact the manufacturer for help.

Data Sheet #2 (for use when $\sigma_{within} \geq 2/3\ \sigma_{total}$)

Device _____ Analyte _____ Concentration _____

Reagent Source/Lot _____ Calibrator Source/Lot _____

	Run 1	Run 2	Run 3
Date/Operator			
Replicate 1 (x_1)			
Replicate 2 (x_2)			
Replicate 3 (x_3)			
$\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3$			
$\bar{x}_d = \frac{\sum_{i=1}^{3} x_i}{3}$			
$x_1 - \bar{x}_d$			
$(x_1 - \bar{x}_d)^2$			
$x_2 - \bar{x}_d$			
$(x_2 - \bar{x}_d)^2$			
$x_3 - \bar{x}_d$			
$(x_3 - \bar{x}_d)^2$			
$\sum_{1}^{3} (x_i - \bar{x}_d)^2 = (x_1 - \bar{x}_d)^2 + (x_2 - \bar{x}_d)^2 + (x_3 - \bar{x}_d)^2$			
$sd^2_{run} = \frac{\sum_{i=1}^{3} (x_i - \bar{x}_d)^2}{2}$			
$\bar{x}_d - \bar{\bar{x}}$ (see below)			
$(\bar{x}_d - \bar{\bar{x}})^2$			

1. Calculation of Grand Mean $= \bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3}{3} =$ _____

2. Calculation of s_{within} :

$$sd^2_{run,average} = \frac{sd^2_{run\ 1} + sd^2_{run\ 2} + sd^2_{run\ 3}}{3} = \_\_\_\_\_$$

$$s_{within} = \sqrt{sd^2_{run,\ average}} = \_\_\_\_\_$$

3. Calculation of s_{total}

s_{within} = _____ , from above

$$B = \frac{\sum_{d=1}^{3} (\bar{x}_d - \bar{\bar{x}})^2}{2} = \frac{(\bar{x}_1 - \bar{\bar{x}})^2 + (\bar{x}_2 - \bar{\bar{x}})^2 + (\bar{x}_3 - \bar{\bar{x}})^2}{2} = \_\_\_\_\_$$

$s_{total} = \sqrt{\frac{n-1}{n} \bullet s_{within}^2 + B}$, where n = number of replicates in each run.

$s_{total} = \sqrt{\frac{2}{3} \bullet s_{within}^2 + B} =$ _____

4. Verification of Within-Run Precision Claim:

Compare calculated s_{within} to claimed σ_{within}:

If calculated s_{within} < claimed σ_{within}, within-run precision has been demonstrated to be consistent with the manufacturer's claim. Go to paragraph 5, "Verification of Total Precision Claim."

If calculated s_{within} > claimed σ_{within}, note that a user's within-run precision can be larger than the manufacturer's claim and not be statistically different from the claim.

$\nu = 6$, $C = 14.45$

Verification value = $\frac{\sigma_{within} \bullet \sqrt{C}}{\sqrt{\nu}} = \sigma_{within} \bullet 1.552 =$ _____

Compare calculated s_{within} to verification value:

If calculated s_{within} < verification value, within-run precision has been demonstrated to be consistent with the manufacturer's claim.

If calculated s_{within} > verification value, within-run precision has not been demonstrated to be consistent with the manufacturer's claim; contact the manufacturer for help.

5. Verification of Total Precision Claim:

Compare calculated s_{total} to claimed σ_{total}:

If calculated s_{total} < claimed σ_{total}, total precision has been demonstrated to be consistent with the manufacturer's claim. Go to Section 6, "Demonstration of Accuracy."

If calculated s_{total} > claimed σ_{total}, note that a user's total precision can be larger than the manufacturer's claim and not be statistically different from the claim.

Calculate degrees of freedom, T:

$$T = \frac{((n-1)s^2_{within} + (nB))^2}{(\frac{n-1}{D})s^4_{within} + (\frac{n^2B^2}{D-1})},$$

where D = number of days, and
n = number of replicates.

$$T = \frac{(2s^2_{within} + 3B)^2}{0.667s^4_{within} + 4.5B^2} = \_\_\_\_\_$$

C = _____ (obtain from Table 1 for T degrees of freedom)

$$\text{Verification value} = \frac{\sigma_{total} \bullet \sqrt{C}}{\sqrt{T}} = \_\_\_\_\_$$

Compare calculated s_{total} to verification value:

If calculated s_{total} < verification value, total precision has been demonstrated to be consistent with the manufacturer's claim.

If calculated s_{total} > verification value, total precision has not been demonstrated to be consistent with the manufacturer's claim; contact the manufacturer for help.

Appendix B. Example Completed Sample Data Recording Sheets – Precision Experiment

Data Sheet #1 (for use when $\sigma_{within} \leq 2/3\ \sigma_{total}$)
(Assume manufacturer's claim for σ_{within} is 1.0 mg/dL and claim for σ_{total} is 2.0 mg/dL)

Device XYZ Analyzer Analyte Glucose Concentration 140 mg/dL

Reagent Source/Lot 2M032397 Calibrator Source/Lot 2RNC52Y

	Run 1	Run 2	Run 3	Run 4	Run 5
Date/Operator	2/20 TF	2/15 JL	2/16 DB	2/19 PO	2/20 GG
Replicate 1 (x_1)	140	138	143	143	142
Replicate 2 (x_2)	140	139	144	143	143
Replicate 3 (x_3)	140	138	144	143	141
Replicate 4 (x_4)	139	137	144	143	142
$\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$	559	552	575	572	568
$\bar{x}_d = \frac{\sum_{i=1}^{4} x_i}{4}$	139.75	138.00	143.75	143.00	142.00
$x_1 - \bar{x}_d$	0.25	0.00	0.75	0.00	0.00
$(x_1 - \bar{x}_d)^2$	0.0625	0.0000	0.5625	0.0000	0.0000
$x_2 - \bar{x}_d$	0.25	1.00	0.25	0.00	1.00
$(x_2 - \bar{x}_d)^2$	0.0625	1.0000	0.0625	0.0000	1.0000
$x_3 - \bar{x}_d$	0.25	0.00	0.25	0.00	1.00
$(x_3 - \bar{x}_d)^2$	0.0625	0.0000	0.0625	0.0000	1.0000
$x_4 - \bar{x}_d$	-0.75	1.00	0.25	0.00	0.00
$(x_4 - \bar{x}_d)^2$	0.5625	1.0000	0.0625	0.0000	0.0000
$\sum_{1}^{4} (x_i - \bar{x}_d)^2 = (x_1 - \bar{x}_d)^2 + (x_2 - \bar{x}_d)^2 + (x_3 - \bar{x}_d)^2 + (x_4 - \bar{x}_d)^2$	0.7500	2.0000	0.7500	0.0000	2.0000
$sd^2_{run} = \frac{\sum_{i=1}^{4} (x_i - \bar{x}_d)^2}{3}$	0.2500	0.6667	0.2500	0.0000	0.6667
$\bar{x}_d - \bar{\bar{x}}$ (see below)	-1.55	-3.30	2.45	1.70	0.70
$(\bar{x}_d - \bar{\bar{x}})^2$	2.4025	10.8900	6.0025	2.8900	0.4900

1. Calculation of Grand Mean $= \bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4 + \bar{x}_5}{5} = \underline{141.3 \text{ mg/dL}}$

2. Calculation of s_{within} :

$$sd^2_{run,average} = \frac{sd^2_{run\ 1} + sd^2_{run\ 2} + sd^2_{run\ 3} + sd^2_{run\ 4} + sd^2_{run\ 5}}{5} = \underline{0.3667\ (mg/dL)^2}$$

$$s_{within} = \sqrt{sd^2_{run,\ average}} = \underline{0.61 \text{ mg/dL}}$$

3. Calculation of s_{total}

s_{within} = 0.61 mg/dL, from above

$$B = \frac{\sum_{d=1}^{5} (\bar{x}_d - \bar{\bar{x}})^2}{4} = \frac{(\bar{x}_1 - \bar{\bar{x}})^2 + (\bar{x}_2 - \bar{\bar{x}})^2 + (\bar{x}_3 - \bar{\bar{x}})^2 + (\bar{x}_4 - \bar{\bar{x}})^2 + (\bar{x}_5 - \bar{\bar{x}})^2}{4} = \underline{5.6688\ (mg/dL)^2}$$

$$s_{total} = \sqrt{\frac{n-1}{n} \bullet s^2_{within} + B}$$, where n = number of replicates in each run.

$$s_{total} = \sqrt{\frac{3}{4} \bullet s^2_{within} + B} = \underline{2.44 \text{ mg/dL}}$$

4. Verification of Within-Run Precision Claim:

Compare calculated s_{within} to claimed σ_{within}:

If calculated s_{within} < claimed σ_{within}, within-run precision has been demonstrated to be consistent with the manufacturer's claim. Go to paragraph 5, "Verification of Total Precision Claim."

If calculated s_{within} > claimed σ_{within}, note that a user's within-run precision can be larger than the manufacturer's claim and not be statistically different from the claim.

Calculated s_{within} (<u>0.60 mg/dL</u>) < claimed σ_{within} (<u>1.00 mg/dL</u>); within-run precision has been demonstrated to be consistent with the manufacturer's claim. (The rest of this paragraph is presented as an example; the user would normally go to paragraph 5, below)

ν = 15, C = 27.49

$$\text{Verification value} = \frac{\sigma_{within} \bullet \sqrt{C}}{\sqrt{\nu}} = \sigma_{within} \bullet 1.354 = \underline{1.354 \text{ mg/dL}}$$

Compare calculated s_{within} to verification value:

If calculated s_{within} < verification value, within-run precision has been demonstrated to be consistent with the manufacturer's claim.

If calculated s_{within} > verification value, within-run precision has not been demonstrated to be consistent with the manufacturer's claim; contact the manufacturer for help.

Calculated s_{within} (0.60 mg/dL) < verification value (1.354 mg/dL); within-run precision has been demonstrated to be consistent with the manufacturer's claim.

5. Verification of Total Precision Claim:

Compare calculated s_{total} to claimed σ_{total}:

If calculated s_{total} < claimed σ_{total}, total precision has been demonstrated to be consistent with the manufacturer's claim. Go to Section 6, "Demonstration of Accuracy."

If calculated s_{total} > claimed σ_{total}, note that a user's total precision can be larger than the manufacturer's claim and not be statistically different from the claim.

Calculated s_{total} (2.44 mg/dL) > claimed σ_{total} (2.00 mg/dL); total precision has not been demonstrated to be consistent with the manufacturer's claim. It is necessary to calculate the verification value and compare calculated s_{total} to the verification value.

Calculate degrees of freedom, T:

$$T = \frac{((n-1)s^2_{within} + (nB))^2}{(\frac{n-1}{D})s^4_{within} + (\frac{n^2B^2}{D-1})},$$

where D = number of days, and
n = number of replicates.

$$T = \frac{(3s^2_{within} + 4B)^2}{0.6s^4_{within} + 4B^2} = \frac{(3 \bullet 0.6^2 + 4 \bullet 5.67)^2}{0.6 \bullet 0.6^4 + 4 \bullet 5.67^2} = \frac{(1.08 + 22.68)^2}{0.0776 + 128.5956} = \frac{564.5376}{128.6732} = \underline{4.39}$$

C = 11.14 (obtain from Table 1 for 4 degrees of freedom)

$$\text{Verification value} = \frac{\sigma_{total} \bullet \sqrt{C}}{\sqrt{T}} = \frac{2.0 \bullet \sqrt{11.14}}{\sqrt{4.39}} = \frac{3.19 \text{ mg/dL}}{}$$

Compare calculated s_{total} to verification value:

If calculated s_{total} < verification value, total precision has been demonstrated to be consistent with the manufacturer's claim.

If calculated s_{total} > verification value, total precision has not been demonstrated to be consistent with the manufacturer's claim; contact the manufacturer for help.

Calculated s_{total} (2.44 mg/dL) < verification value (3.19 mg/dL); total precision has been demonstrated to be consistent with the manufacturer's claim.

Data Sheet #2 (for use when $\sigma_{within} \geq 2/3\ \sigma_{total}$)

(Assume manufacturer's claim for σ_{within} is 1.75 mg/dL and claim for σ_{total} is 2.00 mg/dL)

Device ABC Analyte Glucose Concentration 140 mg/dL

Reagent Source/Lot 2M00567 Calibrator Source/Lot RNC070798

	Run 1	Run 2	Run 3
Date/Operator	3/26 SR	3/27 KW	3/28 DH
Replicate 1 (x_1)	142	140	140
Replicate 2 (x_2)	140	138	139
Replicate 3 (x_3)	140	139	145
$\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3$	422.12	417.11	424.11
$\bar{x}_d = \frac{\sum_{i=1}^{3} x_i}{3}$	140.67	139.00	141.33
$x_1 - \bar{x}_d$	1.33	1.00	-1.33
$(x_1 - \bar{x}_d)^2$	1.7689	1.0000	1.7689
$x_2 - \bar{x}_d$	-0.67	-1.00	-2.33
$(x_2 - \bar{x}_d)^2$	0.4489	1.0000	5.4289
$x_3 - \bar{x}_d$	-0.67	0.00	3.67
$(x_3 - \bar{x}_d)^2$	0.4489	0.0000	13.4689
$\sum_{1}^{3} (x_i - \bar{x}_d)^2 = (x_1 - \bar{x}_d)^2 + (x_2 - \bar{x}_d)^2 + (x_3 - \bar{x}_d)^2$	2.6667	2.0000	20.6667
$sd^2_{run} = \frac{\sum_{i=1}^{3} (x_i - \bar{x}_d)^2}{2}$	1.3334	1.0000	10.3334
$\bar{x}_d - \bar{\bar{x}}$ (see below)	0.34	1.33	1.00
$(\bar{x}_d - \bar{\bar{x}})^2$	0.1156	1.7689	1.0000

1. Calculation of Grand Mean $= \bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3}{3} = 140.33$ mg/dL

2. Calculation of s_{within}

$$sd^2_{run,average} = \frac{sd^2_{run\,1} + sd^2_{run\,2} + sd^2_{run\,3}}{3} = \frac{1.3334 + 1.0000 + 10.3334}{3} = 4.2223$$

$$s_{within} = \sqrt{sd^2_{run,\,average}} = 2.06 \text{ mg/dL}$$

3. Calculation of s_{total}

$s_{within} = 2.06$ mg/dL, from above

$$B = \frac{\sum_{d=1}^{3}(\bar{x}_d - \bar{\bar{x}})^2}{2} = \frac{(\bar{x}_1 - \bar{\bar{x}})^2 + (\bar{x}_2 - \bar{\bar{x}})^2 + (\bar{x}_3 - \bar{\bar{x}})^2}{2} = \frac{2.8845}{2} = 1.4422$$

$s_{total} = \sqrt{\frac{n-1}{n} \bullet s_{within}^2 + B}$, where n = number of replicates in each run.

$$s_{total} = \sqrt{\frac{2}{3} \bullet s_{within}^2 + B} = \sqrt{\frac{2}{3} \bullet 2.06^2 + 1.4422} = 2.07 \text{ mg/dL}$$

4. Verification of Within-Run Precision Claim

Compare calculated s_{within} to claimed σ_{within}:

If calculated s_{within} < claimed σ_{within}, within-run precision has been demonstrated to be consistent with the manufacturer's claim. Go to paragraph 5, "Verification of Total Precision Claim."

If calculated s_{within} > claimed σ_{within}, note that a user's within-run precision can be larger than the manufacturer's claim and not be statistically different from the claim.

Calculated s_{within} (2.06 mg/dL) > claimed σ_{within} (2.00 mg/dL; within-run precision has not been demonstrated to be consistent with the manufacturer's claim. It is necessary to calculate the verification value and compare calculated s_{within} to the verification value.

$\nu = 6$, $C = 14.45$

$$\text{Verification value} = \frac{\sigma_{within} \bullet \sqrt{C}}{\sqrt{\nu}} = \sigma_{within} \bullet 1.552 = 2.72$$

Compare calculated s_{within} to verification value:

If calculated s_{within} < verification value, within-run precision has been demonstrated to be consistent with the manufacturer's claim.

If calculated s_{within} > verification value, within-run precision has not been demonstrated to be consistent with the manufacturer's claim; contact the manufacturer for help.

Calculated s_{within} (2.06 mg/dL) < verification value (2.72); within-run precision has been demonstrated to be consistent with manufacturer's claim.

5. Verification of Total Precision Claim

Compare calculated s_{total} to claimed σ_{total}:

If calculated s_{total} < claimed σ_{total}, total precision has been demonstrated to be consistent with the manufacturer's claim. Go to Section 6, "Demonstration of Accuracy."

If calculated s_{total} > claimed σ_{total}, note that a user's total precision can be larger than the manufacturer's claim and not be statistically different from the claim.

Calculated s_{total} (2.07 mg/dL) > claimed σ_{total} (2.00 mg/dL); total precision has not been demonstrated to be consistent with the manufacturer's claim. It is necessary to calculate the verification value and compare calculated s_{total} to the verification value.

Calculate degrees of freedom, T:

$$T = \frac{((n-1)s^2_{within} + (nB))^2}{(\frac{n-1}{D})s^4_{within} + (\frac{n^2B^2}{D-1})},$$

where D = number of days, and
n = number of replicates.

$$T = \frac{(2s^2_{within} + 3B)^2}{0.667s^4_{within} + 4.5B^2} = \frac{(2 \bullet 2.06^2 + 3 \bullet 1.4422)^2}{0.667 \bullet 2.06^4 + 4.5 \bullet 1.4422^2} = \frac{(8.4872 + 4.3266)^2}{12.0114 + 9.3597} = \frac{164.1934}{21.3711} =$$

$$T = 7.68$$

C = 16.01 (obtain from Table 1 for T degrees of freedom)

$$\text{Verification value} = \frac{\sigma_{total} \bullet \sqrt{C}}{\sqrt{T}} = \frac{2.00\sqrt{16.01}}{\sqrt{7.68}} = 2.89$$

Compare calculated s_{total} to verification value:

If calculated s_{total} < verification value, total precision has been demonstrated to be consistent with the manufacturer's claim.

If calculated s_{total} > verification value, total precision has not been demonstrated to be consistent with the manufacturer's the claim; contact the manufacturer for help.

Calculated s_{total} (2.07 mg/dL) < verification value (2.89 mg/dL); total precision has been demonstrated to be consistent with the manufacturer's claim.

Appendix C. Additional Statistical Explanations and Considerations — Precision Experiment

The appropriate experimental design to employ for demonstrating that a user's calculated precision is consistent with a manufacturer's precision claims depends upon the following four factors:

(a) the statistical false rejection (Type I or "alpha" error) rate;

(b) the statistical false acceptance (Type II or "beta" error) rate;

(c) the number of analyte levels for which imprecision claims are being verified; and

(d) the size of the claimed within-run precision, σ_{within}, relative to the claimed total precision, σ_{total}.

Changes in each of these four factors may affect either the number of days duration of the experiment or the number of replicates per run. In both experimental designs described later in this section, one run per experimental day is required. Consequently, the total number of runs is equal to the number of days duration of the experiment. It is assumed that claims are being verified for all analyte levels and that the same design is used for all levels.

The statistical false rejection testing rate is the chance that the user will falsely conclude that the manufacturer's claims are incorrect. Typical values selected for this error rate are 1% and 5%. Both of the recommended experimental designs assume a false rejection rate of 5%. As the false rejection rate decreases, the required amount of collected data increases. This additional amount of information could result in requirements for more replicates per run or a longer experiment (that is, more days), or both.

The statistical false acceptance testing rate is the chance that the user will falsely conclude that the manufacturer's claims are correct. Typical values selected for this error rate are 1%, 5%, and 10%. As the false acceptance rate decreases, the required amount of collected data increases. False acceptance rates are reported below for both of the recommended experimental designs. If multiple runs per day are performed with either of these experimental designs, then the statistical false acceptance rate will be smaller than the reported rates.

The number of analyte levels being tested indirectly impacts the experimental design through the false rejection rate. Without the proper mathematical corrections, as the number of analyte levels increases, so does the overall false rejection rate. For example, if a 5% false rejection rate is utilized for each level, when four analyte levels are considered, the chance of falsely rejecting at least one claim exceeds 18%. Mathematically, an overall false rejection rate of a specified size, such as the 5% rate associated with both recommended experimental designs, is achieved by lowering the false rejection rate for each level. Unfortunately, lowering the single-level false rejection rate increases the required amount of data to detect a difference between the observed precision and the manufacturer's claimed precision with the same degree of power.

The size of σ_{within} relative to σ_{total} affects both the required number of replicates per run and the duration of the experiment. If σ_{within} is relatively small when compared to σ_{total}, then the number of replicates per run is less important than the total number of runs. In contrast, if σ_{within} is relatively large when compared to σ_{total}, then the number of replicates per run is relatively more important than the experimental duration. If σ_{within} and σ_{total} are approximately equal, then the number of replicates per run and the number of runs are about of equal importance.

This experiment has a false rejection rate of 5% if precision is tested at two concentrations. If precision is tested at more concentrations, the verification value must be adjusted. The effects of increasing the number of concentrations on the performance characteristics of the experiment are discussed below:

If $\sigma_{within} < 2/3\ \sigma_{total}$, and the false rejection rate is set to 5%, and

- two levels are tested, the experiment has an individual false acceptance rate of less than 5% for an increase of σ_{within} to twice its claimed value. For σ_{total} the false acceptance rate is less than 10% for an increase of σ_{total} three times its claimed value.

- three levels are tested, the experiment has an individual false acceptance rate of less than 5% for an increase of σ_{within} to twice its claimed value. For σ_{total} the false acceptance rate is less than 10% for an increase of σ_{total} three times its claimed value.

- four levels are tested, the experiment has an individual false acceptance rate of less than 6% for an increase of σ_{within} to twice its claimed value. For σ_{total} the false acceptance rate is less than 10% for an increase of σ_{total} three times its claimed value.

If $\sigma_{within} > 2/3\ \sigma_{total}$, and the false rejection rate is set to 5%, and

- two levels are tested, the experiment has an individual false acceptance rate of less than 5% for an increase of σ_{within} to three times its claimed value. For σ_{total} the false acceptance rate is less than 15% for an increase of σ_{total} three times its claimed value.

- three levels are tested, the experiment has an individual false acceptance rate of less than 6% for an increase of σ_{within} to three times its claimed value. For σ_{total} the false acceptance rate is less than 15% for an increase of σ_{total} three times its claimed value.

- four levels are tested, the experiment has an individual false acceptance rate of less than 7% for an increase of σ_{within} to three times its claimed value. For σ_{total} the false acceptance rate is less than 15% for an increase of σ_{total} three times its claimed value.

These statements show clearly that this protocol assumes that the user can readily achieve the performance claimed by the manufacturer. It can only detect relatively large deviations from claims. This further emphasizes that this protocol should be applied only in situations in which the performance of the method is already very well characterized.

Appendix D. Sample Data Recording Sheets — Comparison of Patient-Specimens Experiment

Test Method Result	Comp. Method Result	b_i	$b_i - \bar{b}$	$\left(b_i - \bar{b}\right)^2$	$\%b_i$	$\%b_i - \overline{\%b}$	$\left(\%b_i - \overline{\%b}\right)^2$
Sums							

1. Calculation of mean bias in reportable units and percent between the two methods:

$$\bar{b} = \frac{\sum_{i}^{I} b_i}{n} = \_\_\_\_\_$$

$$\overline{\%b} = \frac{\sum_{i}^{I} \%b_i}{n} = \_\_\_\_\_$$

2. Calculation of standard deviations of the bias and percent bias:

$$s_{\bar{b}} = \sqrt{\frac{\sum_{i}^{I} (b_i - \bar{b})^2}{n-1}} = \_\_\_\_\_$$

$$s_{\overline{\%b}} = \sqrt{\frac{\sum_{i}^{I} (\%b_i - \overline{\%b})^2}{n-1}} = \_\_\_\_\_$$

3. Calculation of verification value for bias in reportable units and percent bias:

Compare calculated bias ($\bar{b}$) or percent bias ($\overline{\%b}$) to the claimed bias (β) or the claimed percent bias ($\bar{\beta}$).

If calculated bias ($\bar{b}$) or percent bias ($\overline{\%b}$) < claimed bias (β) or claimed percent bias ($\bar{\beta}$), bias or percent bias has been demonstrated to be consistent with the manufacturer's claim.

If calculated bias ($\bar{b}$) or percent bias ($\overline{\%b}$) > claimed bias (β) or claimed percent bias ($\bar{\beta}$), note that a user's bias or percent can be greater than the manufacturer's claim and not be statistically different from the claim.

Verification value (bias)= $\frac{t \bullet s_{\bar{b}}}{\sqrt{n}} + \beta$ _____

Verification value (percent Bias)= $\frac{t \bullet s_{\overline{\%b}}}{\sqrt{n}} + \beta$ _____

Compare calculated bias or percent bias to appropriate verification value. If calculated bias or percent bias is less than appropriate verification value, bias has been demonstrated to be consistent with the manufacturer's claim. If bias or percent bias exceeds appropriate verification value, bias has not been demonstrated to be consistent with the manufacturer's claim; contact the manufacturer.

Appendix E. Example Completed Sample Data Recording Sheets — Comparison of Patient-Specimens Experiment

Assume the manufacturer's claim for bias is 2.0 mg/dL for a glucose method.

Test Method Result	Comp. Method Result	b_i	$b_i - \bar{b}$	$(b_i - \bar{b})^2$	$\%b_i$	$\%b_i - \overline{\%b}$	$(\%b_i - \overline{\%b})^2$
76	77	-1	-3.5	12.25	-1.30	-3.66	13.39
127	121	6	3.5	12.25	4.96	2.60	6.75
256	262	-6	-8.5	72.25	-2.29	-4.65	21.63
303	294	9	6.5	42.25	3.06	0.70	0.49
29	25	4	1.5	2.25	16.00	13.64	186.02
345	348	-3	-5.5	30.25	-0.86	-3.22	10.39
42	41	1	-1.5	2.25	2.44	0.08	0.01
154	154	0	-2.5	6.25	0.00	-2.36	5.57
398	388	10	7.5	56.25	2.58	0.22	0.05
93	92	1	-1.5	2.25	1.09	-1.27	1.62
240	239	1	-1.5	2.25	0.42	-1.94	3.77
72	69	3	0.5	0.25	4.35	1.99	3.95
312	308	4	1.5	2.25	1.30	-1.06	1.13
99	101	-2	-4.5	20.25	-1.98	-4.34	18.85
375	375	0	-2.5	6.25	0.00	-2.36	5.57
168	162	6	3.5	12.25	3.70	1.34	1.80
59	54	5	2.5	6.25	9.26	6.90	47.59
183	185	-2	-4.5	20.25	-1.08	-3.44	11.85
213	204	9	6.5	42.25	4.41	2.05	4.21
436	431	5	2.5	6.25	1.16	-1.20	1.44
Sums		50		357	47.22		117.61

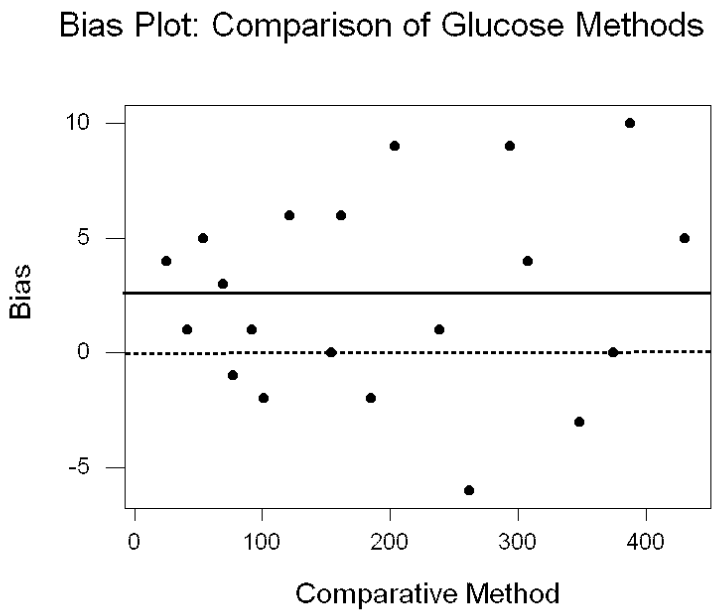

Figure E1. Bias Plot of Sample Data from Comparison of Patient Specimens Experiment. The bias is indicated by the solid line at 2.05 mg/dL. The ideal bias (zero) is indicated by the dashed line at zero. Note that the bias is approximately constant over the concentration range of the experiment.

1. Calculation of mean bias in reportable units and percent between the two methods:

$$\bar{b} = \frac{\sum_{i}^{I} b_i}{n} = \frac{50}{20} = 2.50 \text{ mg/dL}$$

$$\overline{\%b} = \frac{\sum_{i}^{I} \%b_i}{n} = \frac{47.22}{20} = 2.36\%$$

2. Calculation of standard deviations of the bias and percent bias:

$$s_{\bar{b}} = \sqrt{\frac{\sum_{i}^{I} (b_i - \bar{b})^2}{n-1}} = \sqrt{\frac{357.00}{19}} = 4.33 \text{ mg/dL}$$

$$s_{\overline{\%b}} = \sqrt{\frac{\sum_{i}^{I} (\%b_i - \overline{\%b})^2}{n-1}} = \sqrt{\frac{346.08}{19}} = 4.27\%$$

3. Calculation of verification value for bias in reportable units and percent bias:

Compare calculated bias ($\bar{b}$) or percent bias ($\overline{\%b}$) to the claimed bias (β) or the claimed percent bias ($\bar{\beta}$).

If calculated bias ($\bar{b}$) or percent bias ($\overline{\%b}$) < claimed bias (β) or claimed percent bias ($\bar{\beta}$), bias or percent bias has been demonstrated to be consistent with the manufacturer's claim.

If calculated bias ($\bar{b}$) or percent bias ($\overline{\%b}$) > claimed bias (β) or claimed percent bias ($\bar{\beta}$), note that a user's bias or percent can be greater than the manufacturer's claim and not be statistically different from the claim.

Calculated bias (2.50 mg/dL) > claimed bias (2.00 mg/dL); note that a user's bias can be greater than the manufacturer's claim and not be statistically different from the claim. It is necessary to calculate the verification value and compare calculated bias to the verification value.

$$\text{Verification value (bias)} = \frac{t \bullet s_{\bar{b}}}{\sqrt{n}} = \frac{2.539 \bullet 4.33 \text{ mg/dL}}{4.47} + 2.0 \text{ mg/dL} = 4.46 \text{ mg/dL}$$

The calculated bias (2.50 mg/dL) is less than the verification value (4.46 mg/dL); bias has been shown to be consistent with the manufacturer's claim.

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

EP15-P: *User Demonstration of Performance for Precision and Accuracy; Proposed Guideline*

General

1. Generally the document is fine and presented in a fashion that will allow it to be followed easily in the laboratory. I found the examples in the Appendixes very useful.

   There is one point, however, that requires attention. The document is intended for use when a method is being set up and should be applicable to the "widest range of analytes and device complexity." The experimental design requires the measurement of samples over a maximum of five days.

   Most modern analyzers use calibrator materials for setting up assays. These calibrators may need to be run daily or less frequently. We have found that often the shift due to recalibration is greater than the shift in running the QC. In the method described in this proposed guideline, there is a possibility that the experimental work could be carried out without an analyzer recalibration. The data generated could then indicate that the method performance is better than its actual performance over a long period of time.

   My suggestion is it needs to be stated somewhere in the document that during the five-day measurement period, instrument calibration should be performed at least daily. This would give a true indication of the method performance.

- **The subcommittee agrees with this comment. A change in Section 5.2 (5) has been made as follows: "Calibrate as specified in the manufacturer's instructions for operators. If the manufacturer indicates in its claim that its precision data were generated over multiple calibration cycles, then the operator may choose to recalibrate during the experiment."**

Section 4.2

2. As a leading manufacturer for control products, we are concerned about the emphasis placed on manufacturer recommendations in Section 4.2. We believe credible anecdotal evidence exists that suggests manufacturer's control products are not always "unbiased." We would suggest the following alternative wording for the third sentence: "Often the easiest way to demonstrate this during an evaluation is to use assayed control materials that are well characterized by the laboratory or by peer-group analysis."

- **To address these concerns, the subcommittee has added or changed the text as follows: in Section 4.1, last sentence changed to: "Training should include the use of actual sample material (including pools, controls, leftover patient specimens, or any other test materials appropriate for the device)." In Section 4.2, last sentence changed to read: "To demonstrate this during an evaluation, use the control procedures recommended by the manufacturer." In addition, the following text has been added to the end of Section 4.2: "Due to the short duration of this protocol, the estimated standard deviations should not be used by themselves to establish quality control ranges. For guidance on establishing ongoing quality control procedures, refer to the most current edition of NCCLS document C24—*Statistical Quality Control for***

***Quantitative Measurements: Principles and Definitions.*"** **This reference has also been added to the section on Related NCCLS Publications at the end of the document.**

Section 6

3. (Section 6, Demonstration of Accuracy, page 10): Please consider adding the following sentence: "Manufacturer's instructions should be followed for collection and handling of patients' specimens to be assayed on the manufacturer's instrument."

- **The subcommittee agrees with this comment. The following text has been added to Section 6.1 after the fourth sentence in Section 6.1 (1*):* "Follow the manufacturer's instructions for collection and handling of patients' specimens to be assayed on the test system."**

Summary of Delegate Comments and Subcommittee Responses

EP15-A: *User Demonstration of Performance for Precision and Accuracy; Approved Guideline*

General

1. This is excellent guidance for assays that have reportable ranges. I would like to see similar guidance for tests that have no reportable range, but produce a "positive" or "negative" result only.

- **The subcommittee appreciates the compliment and will forward the commenter's suggestion for development of a guideline for qualitative result reporting to the appropriate area committee.**

2. A few years ago, NCCLS decided to accept the Secretariat of ISO TC 212, Clinical Laboratory Testing and *in vitro* Diagnostic Test Systems. This fact should have consequences on the terminology used in NCCLS standards. We think that NCCLS should incorporate the metrological concepts and terms recommended by ISO (and endorsed by IFCC and IUPAC). For this reason, we have made corrections on the draft in order that this document is consistent with ISO metrological terminology for clinical laboratory science.

- **NCCLS recognizes that harmonization of definitions facilitates global application of standards, and will remain steadfast in its commitment to global harmonization of nomenclature by employing terminology in its documents that is generally used internationally. This initiative includes a mechanism to resolve ISO/CEN/ NCCLS differences in nomenclature.**

  However, NCCLS is also aware that legally required use of terms, regional usage, and different consensus timelines are all obstacles; therefore, implementation of this policy must be an evolutionary and educational process that begins with new projects and revisions of existing documents. This document is scheduled for revision in 2004; at that time, all definitions will be reviewed for consistency with international use, and revised appropriately. More specifically, the users of EP15-A should understand that the fundamental meanings of the terms (e.g., *accuracy* vs. *trueness; analytical method* vs. *measurement procedure; analyte* vs. *quantity*) are identical, and to facilitate understanding, all equivalent terms are included in the guideline's Definitions section.

3. One overall comment. These methods described here are extremely sensitive to the assumption of Normality, although this is not mentioned in the text. Essentially, the observations need to be normally distributed around their respective within-run means. Also, the day-to-day effects need to be normally distributed around the grand mean of the experiment.

- **The subcommittee appreciates the comment, yet believes that incorporating the concept of Normality will not aid in clarifying the intended use of the guideline.**

4. The mathematics is very good to see, though the work sheets remind me of Schedule D of the IRS form 1040. There isn't much you can do about this and still give you something that nonmathematician's can use.

- **The approach described in this protocol measures the precision and bias of the method and compares the measured quantities to the manufacturer's claims. Inevitably, some mathematics is required in this process; however, the subcommittee has made attempts to minimize the mathematics required to make reliable decisions about the observed performance for precision and accuracy.**

5. I believe there is little benefit to attempting to demonstrate that some sources of variability can be reproduced when the power of the testing is limited and confidence in the conclusion is questionable. For example, evaluating the reproducibility of calibration or reagent lots (with approximately three calibrations—and lot variability is not mentioned) is not justifiable, considering the scope of the project.

- **As recommended in Section 5.2 (5), the user *may* perform multiple calibrations if the manufacturer has included them in the precision claims. It is not the intent of this guideline to require multiple calibrations.**

6. The protocol should be restructured to clearly state that manufacturer's claims for within-run and total within-lab precision (text uses total precision) may be verified by a five-day precision test, 2 runs per day.

- **The subcommittee understands this comment to suggest the use of "total within-lab precision" instead of simply "total precision." The subcommittee believes this has been accomplished in Paragraph 2.2.**

7. Including multiple calibration events within this five-day evaluation period is not appropriate.

- **As stated above, it is not the intent of this guideline to require multiple calibrations.**

8. Comparison of patient specimen results to another is logistically difficult and very expensive. Issues can also arise about qualifying the comparative method, which is beyond the scope or intention of these demonstrations. Therefore, patient comparison experiments are a valid, but impractical, second choice.

- **The subcommittee chose bias calculated from the patient sample comparison of methods experiment as the primary method for demonstrating accuracy to align with companion document EP9—*Method Comparison and Bias Estimation Using Patient Samples*, which recommends stating bias as the outcome of the comparison of methods experiment. This guideline recommends that the user compare to the same method employed as the comparative method by the manufacturer, in order for bias to have any meaning. The qualification of the comparative method is not in question here; it is the one chosen by the manufacturer for his claim.**

9. The protocol should clearly state that a calibration verification step should be completed before accuracy is assessed. If calibration is verified, then, the protocol should be completed using a single calibration.

- **The subcommittee agrees that the accuracy protocol should mention calibration, however, calibration verification is not necessary. For many analytes, the choice of materials for calibration verification is limited. Text added to Section 4 suggests calibration should be verified during this period, if appropriate.**

Foreword

10. Evaluation, as defined here, is not consistent with standard English: the act of evaluation (the definition of "-tion" is the act of).

- **The term "evaluation" has been redefined as follows, "a measurement of the analytical performance characteristics of a new method by means of laboratory experiments."**

11. Paragraph 3. Consistency with peer group is only a possible outcome if PT materials are used for demonstrating accuracy.

- **The first sentence of this paragraph has been revised to read, "This guideline has been developed to guide the user through minimum studies necessary to demonstrate that the user can obtain precision and accuracy performance consistent with the manufacturer's claims, and, if proficiency testing materials are used, consistent with the test system's peer group, as well."**

Introduction

12. The use of "within-run" and "total" precision is obsolete. The internationally accepted terms, which appear to be very understandable to all users, are "repeatability" and "reproducibility" respectively.

- **The terms "within-run precision" and "total precision" have been maintained for consistency with companion document EP5, *Evaluation of Precision Performance of Clinical Chemistry Devices*, but the definitions section has been revised and compares the definitions for repeatability and reproducibility to "within-run precision" and "total precision." See also Comment No. 2 and the Note on Terminology in the Foreword.**

Section 1.1

13. Paragraph 5. "Peer group" might not be obvious to the audience.

- **As recommended, the term "Peer group" has been added to Section 4 Definitions.**

14. Paragraph 7. No protocol will "make" a product's claims meet the goal. The intent, I think, is to use more rigorous protocols to obtain higher confidence in the findings.

- **The text has been modified for clarity. The sentence now reads, "Other more rigorous NCCLS protocols should be employed to validate the method's performance against the user's needs."**

Section 2.3

15. Paragraph 1. Reverse the options in Section 6, switch options (a) and (b) to demonstrate preference for ease of use reasons.

- **The order in which the options for an accuracy evaluation experiment are presented in this guideline will be maintained to ensure consistency with companion document EP5—*Evaluation of Precision Performance of Clinical Chemistry Devices.***

16. Section 2.3(b). PT materials are not the only assayed materials that can be used to demonstrate accuracy.

- **As stated in Section 2.3(b), "Accuracy may be assessed by analyzing proficiency test materials and other assayed materials." A description of acceptable assayed materials is provided in Section 6.2, "Recovery of Expected Values from Assayed Reference Materials."**

17. "Measure," (not "analyze") is the accepted term for these actions.

- **The subcommittee appreciates this comment. The subcommittee reached consensus on language used in this section; therefore, the term "analyze" will be maintained so as to be consistent with EP5—*Evaluation of Precision Performance of Clinical Chemistry Devices.***

Section 3

18. I agree with the VIM/ISO co-definitions. However for "total error," I recommend removing the words "maximum permissible." The use of "total error" is a measured error, rather than a goal.

- **The text has been modified as recommended.**

Section 4.3

19. In the last sentence, material should be "as close as possible to human whole blood" but should not necessarily preclude the use of animal cells and for plasma.

- **The text does not preclude the use of animal materials as long as they are "as close as possible to human whole blood." The key point is that the material should mimic a patient specimen as closely as possible.**

20. Paragraph 1, line 2. Precision might not be related to concentration, and it is not the justification to measure performance at the medical decision levels.

- **The text has been modified to read, "... analyte concentrations should be focused at or near medical decision points."**

21. Paragraph 1. Not all products come with a package insert. It is common for instructions to be sent separately for many systems.

- **The text has been modified to read "(This information is in the package insert or instructions for use provided by the manufacturer of the test system under evaluation.)"**

Section 5.1

22. Paragraph 2. False rejection is not a common term for this audience.

- **A parenthetical statement has been added to refer the user to Appendix C for detailed information on statistical false rejection rate.**

23. Paragraph 2 …maximum of two analyte levels ...There could be some misleading estimation of variances if two levels are very different yet their CVs are constant, the higher level will be underestimated and the lower level will be overestimated. One of the proper approaches could be restricting two levels to be variation compatible or homogeneous, within-run and between-run. If there is no information about the variation compatibility, the estimation should be performed by level.

- **Although this is a theoretical possibility, the designs should, in most cases, be the same as they are based on the ratio of the within-run variance to the total variance, and not the within-run variance to between-run variance. The design provides quite a bit of leverage in the values that the within-run and between-run variances can assume and does not assume that the within-run and the between-run variances are the same across levels.**

  For example, at level 1 the within-run variance could be 55% of the total and the between-run variance could be 45% of the total; while at level 2 the within-run variance could be 20% of the total and the between-run variance could be 80% of the total. In either case the design would be the same since the $\sigma_{\text{within-run}} < 2/3\ \sigma_{\text{total}}$.

24. Paragraph 2. What will the 5% rejection rate and the sample sizes stated yield (i.e., what would be the resultant power or accuracy)? From a statistical point of view, this is an incomplete procedure without indicating resultant power or accuracy. The number of days and number of runs per day depend upon between-day and within-day variances. A "one-size-fits-all" approach may be misleading when drawing conclusions.

- **The power of the test for the different experimental designs suggested in this guideline is discussed in Appendix C. While it is true that more replicates could improve the estimation of the within variability, the design chosen does have a power of 90%. While experimental designs for every combination of between and within day variances could be derived, the design suggested is optimal for when $\sigma_{\text{within-run}} > 2/3\ \sigma_{\text{total}}$ and $\sigma_{\text{within-run}} < 2/3\ \sigma_{\text{total}}$.**

  The objective of this guideline is to demonstrate, using a minimal design, that the precision and accuracy of a method are consistent with the claims of the manufacturer. Further conclusions relative to the precision or accuracy of the method, based on this minimal design should not be drawn. As stated in the Foreword, it is assumed that the method's performance has been evaluated using protocols designed to validate and verify performance.

25. Paragraph 2 (b)"With $\sigma_{\text{within}} > 2/3\ \sigma_{\text{total}}$ ..." It is understood that fewer days would be needed because the between-run variation decreased, but why fewer replicates per day as well while the within-variation increased? More replicates per run could improve the estimation of within variation when this variation was known to be large. I could not understand why the number of replicates is reduced. It would be nice to provide reference(s) addressed how these results were concluded.

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

Section 5.4

26. It would be useful to indicate in this section that the analysis is performed separately for each concentration level. It becomes obvious later when working through the statistical reasoning in the Appendix, but it bears noting in the beginning of this section.

- **The text in the last sentence has been modified to read: "Separate calculations should be performed for each level of concentration."**

Section 5.4.1

27. Equation for S_{within}: the initial limits of summation are missing on both summations. They should be d=1 and i=1.

- **The equation has been modified to provide initial limits of summation as suggested.**

28. Last line. Change "may be incorrect" to "will be incorrect." The S_{within} will be underestimated because the denominator D (n-1) will be larger than the actual degrees of freedom.

- **The text has been modified as suggested.**

Section 5.4.2

29. $$s_{total} = \sqrt{\frac{n-1}{n} \bullet s^2_{within} + B}$$

Why use (n-1)/n s^2_{within}? The estimated within and between variances are unbiased. The adjustment (n-1)/n makes the within variance a biased estimator. It is not quite logical to me that the total variance is the sum of a biased estimator and an unbiased estimator. One can use either the sum of unbiased estimators or sum of biased estimators (i.e., maximum likelihood estimators). But one has to be cautious because n and D are small as the procedure suggested. It could result in different conclusion when divided by n or n-1, and by D or D-1. People use different estimators in different situations but it is odd to take the sum of a biased estimator and an unbiased estimator. If this sum has to be used as the estimation of total variance, references should be provided.

- **The estimates are MLEs, and hence, do not have to be unbiased. The estimations of all of these quantities are identical to those presented in NCCLS document EP-5—*Evaluation of Precision Performance of Clinical Chemistry Devices.***

Section 5.5

30. Paragraph 1 (2) and paragraph 2 (2). The sections using sophisticated statistical terms need more clarity. I am generally familiar with these terms, but I could not understand what "C for these testing points" meant.

- **As stated, the percentage point "C" is a chi-squared value. Users are referred to Table 1 for a list of values for "C."**

31. $1 - \alpha / \ell$

When more than one level is used, the procedure assumes that the within-run and between-run variances are compatible or "same" across levels. This is the reason that the procedure ends up with one estimator for the within-run and one for the between-run. I am not sure what the purpose is for this adjustment. If the assumption of compatible variations is in doubt, this adjustment will not make it correct or compensate anything. The estimation will always be either underestimated or overestimated.

- **See the response to comment 22 above.**

Section 5.6

32. Section 5.6 (2). The parenthetical statement after (2) should read "chi-squared," not "chi-square."

- **This editorial correction has been made to the text.**

Section 6

33. Section 6 (1). This option is presented as too strong a preference. Do not use "should be performed." This must be given as the second option because it is much more expensive, more complicated, and brings up issues about qualifications of the test comparison – very often out of the small lab's control or knowledge.

- **The protocol described in Section 6 "Demonstration of Accuracy" is provided as a recommendation to users. The remainder of the paragraph provides the subcommittee's rationale for the precedence of the experiment.**

34. Section 6 (1), last sentence, I recommend that the wording be changed to read as follows: The comparison of patient specimens can also form the basis of establishing the relationship between multiple methods. This will ensure the laboratory's ability to provide equivalent results irrespective of which method is used to assay a patient specimen (see..)."

- **The text has been modified as suggested.**

Section 6.1

35. Section 6.1, paragraph 2. A discussion as to whether the testing should be done using a single determination or replicates would be of value.

- **The text of paragraph 3 has been modified to address the comment. It reads as follows: "The experimental protocol for comparing results of patient specimen testing in this guideline has been designed as a demonstration protocol. To keep the experimental work simple, each patient specimen is tested singly by the comparative and test methods. Because it has relatively low power to detect bias between methods, it should only be employed when the user anticipates close agreement between the methods. Otherwise, the user should employ NCCLS document EP9, *Method Comparison and Bias Estimation Using Patient Samples*, because it includes more patient specimens and testing in duplicate by both methods."**

36. Section 6.1 (5) and (6). Initial limits of summation are missing on both summations. They should be i=1 in all four equations.

- **The equations have been modified to provide initial limits of summation as suggested.**

37. Paragraph 2, lines 1-3. These sections are inconsistent with the scope of this project. Consistent accuracy with the manufacturer's claim is not a relevant concept. The object is to demonstrate accuracy with a simple procedure so any reference to comparing to the method the manufacturer might have used is irrelevant.

- **Paragraph 3 of the Foreword states that this is the scope of the project.**

38. Section 6.1 (1), line 9. The decision to use stored specimens must primarily be based on the manufacturer's instruction, even for abnormal concentrations.

- **The sentence has been modified to read, "If stored specimens are used, they should be refrigerated, if consistent with analyte stability and manufacturer's instructions, to avoid possible artifacts introduced in the freeze-thaw cycle."**

39. Section 6.1, (1) line 3. Instructions should include the precaution that samples containing known interferences should not be included in the comparison study. The user should refer to both manufacturers' instructions for use.

- **An additional sentence has been added to caution users, after the third sentence that ends "...test methods." "Exclude specimens containing substances identified as interferents (in the manufacturer's instructions) for either the test method or comparative method."**

40. The recommendation in section 6.1(1) that "assays by each method should be completed within four hours of each other in the same day" is rather restrictive for many analytes that are relatively stable when handled properly and not restrictive enough for some (like CO_2). Since analyte instability will tend to cause random deviations in the result comparisons of varying magnitude but in a consistent direction; and since the recommendation I made in this section to examine the assay results after each testing event, this four-hour recommended time window is probably not necessary. I suggest using the manufacturer's strong recommendations for the methods and evaluating for possible aging effects by looking for unexpected, random, undirectional deviations in comparison results. Note that if my comparative method is sending the sample to a reference lab, the four-hour window is not feasible at all.

- **Section 6.1 has been modified to include the manufacturer's recommendation for specimen stability. The recommendation of a four-hour time window is included in order to prevent small specimen stability problems from appearing to be bias between the two methods.**

Section 6.1.1

41. Section 6.1.1 (2). Clarification is needed for (100-α) percent point.

- **The text has been modified to clarify the (100-α) percentage point referred to in (2).**

42. Paragraph 3 (3). This section needs to be revised. Manufacturers generally describe accuracy (or bias) by correlation parameters (usually as slope and intercept), without a value of uncertainty for the correlation.

- **Section 6.1 (paragraphs 4 through 6) provides a description and calculation for bias.**

43. In regards to the Note in this section, what if "bias" is not constant throughout the reportable range?

- **The Note has been changed to read "This method can be used to test the bias only when the bias or percent bias is constant throughout the reportable range, or to test the biases or percent biases calculated by splitting the data as described in Section 6.1 (4)."**

Section 6.2

44. Paragraph 1, line 5. This is a misleading statement that decreases the confidence a reader would have in reference materials. In most cases, the matrix does not affect recovery.

- **The subcommittee agrees with the commentor. The wording has been changed to read, "The difference in matrix *may cause* an altered analytical response which is unique to a particular material-method combination."**

45. Paragraph 1, line 9. The use of assayed controls does not limit their use to peer group evaluation if the values can be shown to be traceable to a reference material or method.

- **The subcommittee believes the wording of the sentence is clear, i.e., the value was assigned specifically for the method of interest. The remainder of the paragraph explains the subcommittee's rationale for this statement. The final sentence recommends that the manufacturer should be consulted for advice regarding appropriate reference materials for validation of accuracy.**

46. Section 6.2 (d). Are there "regional QC programs anymore, now that CAP stopped theirs? Perhaps change "regional" to "interlaboratory" or "manufacturer sponsored."

- **The text has been modified to read: "interlaboratory quality control program."**

Section 6.2.1

47. Section 6.2.1 (1), line 2. Delete reference to simulating proficiency testing. The objective of this guideline is not to assure passing PT.

- **This guideline presents a mechanism by which users can demonstrate the precision and accuracy of a test method following a PT failure. The subcommittee believes it is appropriate to recommend simulating PT as an example.**

Appendix C

48. Add the following text to the end of paragraph 5, "to detect a specified alternative with the same degree of power."

- **The text has been modified to read: "Unfortunately, lowering the single-level false rejection rate increases the required amount of data to detect a difference between the observed precision and the manufacturer's claimed precision with the same degree of power."**

49. Last paragraph. Provide some assistance for how a user can decide if a method is well characterized.

- **The issue of characterization is discussed in the fourth paragraph of the Foreword.**

50. In the next-to-last paragraph, the third sentence gives an example where 10 analyte levels leads to a 40% chance of at least one false rejection. In this document, we only refer to 2,3,4 levels (Table 1, page 9). Perhaps we should change the example to state 4 levels instead of 10, and the chance of false rejection exceeds 18% instead of 40%.

- **The subcommittee agrees. The text has been changed.**

51. Figure E1, it appears that "bias" is not constant over the concentration range, contrary to the statement under the figure. This could be fixed by changing one or two of the data points in the range <100 (and the corresponding calculations in the worksheets).

- **The subcommittee agrees. The data set and corresponding calculations have been revised.**

Related NCCLS Publications*

C24-A2 **Statistical Quality Control for Quantitative Measurements: Principles and Definitions; Approved Guideline—Second Edition (1999).** This guideline provides definitions of analytical intervals; plans for quality control procedures; and guidance for quality control applications.

C28-A2 **How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline (2000).** This document provides guidance for determining reference values and reference intervals for quantitative clinical laboratory tests.

EP5-A **Evaluation of Precision Performance of Clinical Chemistry Devices; Approved Guideline (1999).** This document offers guidelines 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; and manufacturer's guidelines for establishing claims.

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

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

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

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

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

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.

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 1-56238-451-1