EP10-A2
Vol. 22 No. 29
Replaces EP10-A
Vol. 18 No. 6

Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition

This guideline provides experimental design and data analysis for preliminary evaluation of the performance of a measurement procedure or device.

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

NCCLS

NCCLS...

Serving the World's Medical Science Community Through Voluntary Consensus

NCCLS is an international, interdisciplinary, nonprofit, standards-developing, and educational organization that promotes the development and use of voluntary consensus standards and guidelines within the healthcare community. It is recognized worldwide for the application of its unique consensus process in the development of standards and guidelines for patient testing and related healthcare issues. NCCLS is based on the principle that consensus is an effective and cost-effective way to improve patient testing and healthcare services.

In addition to developing and promoting the use of voluntary consensus standards and guidelines, NCCLS provides an open and unbiased forum to address critical issues affecting the quality of patient testing and health care.

PUBLICATIONS

An NCCLS document is published as a standard, guideline, or committee report.

Standard A document developed through the consensus process that clearly identifies specific, essential requirements for materials, methods, or practices for use in an unmodified form. A standard may, in addition, contain discretionary elements, which are clearly identified.

Guideline A document developed through the consensus process describing criteria for a general operating practice, procedure, or material for voluntary use. A guideline may be used as written or modified by the user to fit specific needs.

Report A document that has not been subjected to consensus review and is released by the Board of Directors.

CONSENSUS PROCESS

The NCCLS voluntary consensus process is a protocol establishing formal criteria for:

- the authorization of a project
- the development and open review of documents
- the revision of documents in response to comments by users
- the acceptance of a document as a consensus standard or guideline.

Most NCCLS documents are subject to two levels of consensus—"proposed" and "approved." Depending on the need for field evaluation or data collection, documents may also be made available for review at an intermediate (i.e., "tentative") consensus level.

Proposed An NCCLS consensus document undergoes the first stage of review by the healthcare community as a proposed standard or guideline. The document should receive a wide and thorough technical review, including an overall review of its scope, approach, and utility, and a line-by-line review of its technical and editorial content.

Tentative A tentative standard or guideline is made available for review and comment only when a recommended method has a well-defined need for a field evaluation or when a recommended protocol requires that specific data be collected. It should be reviewed to ensure its utility.

Approved An approved standard or guideline has achieved consensus within the healthcare community. It should be reviewed to assess the utility of the final document, to ensure attainment of consensus (i.e., that comments on earlier versions have been satisfactorily addressed), and to identify the need for additional consensus documents.

NCCLS standards and guidelines represent a consensus opinion on good practices and reflect the substantial agreement by materially affected, competent, and interested parties obtained by following NCCLS's established consensus procedures. Provisions in NCCLS standards and guidelines may be more or less stringent than applicable regulations. Consequently, conformance to this voluntary consensus document does not relieve the user of responsibility for compliance with applicable regulations.

COMMENTS

The comments of users are essential to the consensus process. Anyone may submit a comment, and all comments are addressed, according to the consensus process, by the NCCLS committee that wrote the document. All comments, including those that result in a change to the document when published at the next consensus level and those that do not result in a change, are responded to by the committee in an appendix to the document. Readers are strongly encouraged to comment in any form and at any time on any NCCLS document. Address comments to the NCCLS Executive Offices, 940 West Valley Road, Suite 1400, Wayne, PA 19087, USA.

VOLUNTEER PARTICIPATION

Healthcare professionals in all specialties are urged to volunteer for participation in NCCLS projects. Please contact the NCCLS Executive Offices for additional information on committee participation.

Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition

Abstract

NCCLS document EP10-A2—*Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition* is intended to facilitate a limited, preliminary evaluation of the performance of a measurement procedure or device. Using the experimental design and data analysis procedure described, determination of whether a device has problems that require further evaluation or referral to the manufacturer can be done with a minimum expenditure of time and material. Included in Appendixes A and B, are sample data sheets that should facilitate the analysis of the data. Appendix C contains a more sophisticated, powerful, statistical method for determining the possible causes of imprecision.

NCCLS. *Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition*. NCCLS document EP10-A2 (ISBN 1-56238-482-1). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.

EP10-A2
ISBN 1-56238-482-1
ISSN 0273-3099

Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition

Volume 22 Number 29

Jan S. Krouwer, Ph.D.
Daniel W. Tholen, M.S.
Carl C. Garber, Ph.D.
Henk M.J. Goldschmidt, Ph.D.
Martin Harris Kroll, M.D.
Kristin Linnet, M.D., Ph.D.
Kristen Meier, Ph.D.
Max Robinowitz, M.D.

Suggested Citation

(NCCLS. *Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition.* NCCLS document EP10-A2 [ISBN 1-56238-482-1]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.)

Proposed Guideline
December 1985

Tentative Guideline
June 1989

Tentative Guideline—Second Edition
September 1993

Approved Guideline
May 1998

Approved Guideline—Second Edition
December 2002

ISBN 1-56238-482-1
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**
Carl C. Garber, Ph.D.	Quest Diagnostics Assurance Teterboro, New Jersey
Henk M.J. Goldschmidt, Ph.D.	Tilburg, The Netherlands
Martin Harris Kroll, M.D.	Dallas Veterans Affairs Medical Center Dallas, Texas
Kristian Linnet, M.D., Ph.D.	Psychiatric University Hospital Risskov, Denmark
Kristen Meier, Ph.D.	FDA Center for Devices/Rad. Health Rockville, Maryland
Max Robinowitz, M.D.	FDA Center for Devices/Rad. Health Rockville, Maryland

Advisors

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

Acknowledgements

The Area Committee on Evaluation Protocols would also like to recognize the valuable contributions of the members and advisors of the Subcommittee on Preliminary Evaluation of Quantitative Clinical Laboratory Methods that developed the first approved edition of this guideline.

Jan S. Krouwer, Ph.D.
Kicab Castañeda-Méndez
John M. Dawson
George F. Johnson, Ph.D.
Abba Sharma, Ph.D.

Active Membership
(as of 1 October 2002)

Sustaining Members

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

Professional Members

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

Government Members

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

Industry Members

AB Biodisk
Abbott Laboratories
Abbott Laboratories, MediSense Products
Acrometrix Corporation
Alifax S.P.A.
Ammirati Regulatory Consulting
Anaerobe Systems
Asséssor
AstraZeneca
AstraZeneca R & D Boston
Avant Immunotherapeutics, Inc.
Aventis
Axis-Shield POC AS
Bayer Corporation – Elkhart, IN
Bayer Corporation – Tarrytown, NY
Bayer Corporation – West Haven, CT
Bayer Medical Ltd.

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

Trade Associations

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

Associate Active Members

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

Albemarle Hospital (NC)
Allegheny General Hospital (PA)
Allina Health System (MN)
Alton Ochsner Medical Foundation (LA)
Antwerp University Hospital (Belgium)
Arkansas Department of Health
ARUP at University Hospital (UT)
Armed Forces Research Institute of Medical Science (APO, AP)
Associated Regional & University Pathologists (UT)
Aurora Consolidated Laboratories (WI)
Azienda Ospedale Di Lecco (Italy)
Bay Medical Center (MI)
Baystate Medical Center (MA)
Bbaguas Duzen Laboratories (Turkey)
Bermuda Hospitals Board
Bo Ali Hospital (Iran)
Brooks Air Force Base (TX)
Broward General Medical Center (FL)
Cadham Provincial Laboratory
Calgary Laboratory Services
Carilion Consolidated Laboratory (VA)
Cathay General Hospital (Taiwan)
Central Peninsula General Hospital (AK)
Central Texas Veterans Health Care System
Centre Hospitalier Regional del la Citadelle (Belgium)
Centro Diagnostico Italiano (Milano, Italy)
Champlain Valley Physicians Hospital (NY)
Chang Gung Memorial Hospital (Taiwan)
Changi General Hospital (Singapore)
The Charlotte Hungerford Hospital (CT)
Children's Hospital (LA)
Children's Hospital (NE)
Children's Hospital & Clinics (MN)
Children's Hospital Medical Center (Akron, OH)
Children's Hospital of Philadelphia (PA)
Children's Medical Center of Dallas (TX)
Clarian Health–Methodist Hospital (IN)
Clendo Lab (Puerto Rico)
Clinical Laboratory Partners, LLC (CT)
CLSI Laboratories (PA)
Columbia Regional Hospital (MO)
Commonwealth of Kentucky
Community Hospital of Lancaster (PA)
CompuNet Clinical Laboratories (OH)
Cook County Hospital (IL)
Cook Children's Medical Center (TX)
Covance Central Laboratory Services (IN)
Danish Veterinary Laboratory (Denmark)
Danville Regional Medical Center (VA)
Delaware Public Health Laboratory
Department of Health & Community Services (New Brunswick, Canada)
DesPeres Hospital (MO)
DeTar Hospital (TX)
Detroit Health Department (MI)
Diagnosticos da América S/A (Brazil)
Dr. Everett Chalmers Hospital (New Brunswick, Canada)
Doctors Hospital (Bahamas)
Duke University Medical Center (NC)
Dwight David Eisenhower Army Med. Ctr. (GA)
E.A. Conway Medical Center (LA)
Eastern Maine Medical Center
East Side Clinical Laboratory (RI)
Eastern Health (Vic., Australia)
Elyria Memorial Hospital (OH)
Emory University Hospital (GA)
Esoterix Center for Infectious Disease (TX)
Fairview-University Medical Center (MN)
Federal Medical Center (MN)
Florida Hospital East Orlando
Focus Technologies (CA)
Foothills Hospital (Calgary, AB, Canada)
Fresenius Medical Care/Spectra East (NJ)
Fresno Community Hospital and Medical Center
Frye Regional Medical Center (NC)
Gambro BCT (CO)
Geisinger Medical Center (PA)
Grady Memorial Hospital (GA)
Guthrie Clinic Laboratories (PA)
Hahnemann University Hospital (PA)
Harris Methodist Erath County (TX)
Harris Methodist Fort Worth (TX)
Hartford Hospital (CT)
Headwaters Health Authority (Alberta, Canada)
Health Network Lab (PA)
Health Partners Laboratories (VA)
Heartland Regional Medical Center (MO)
Highlands Regional Medical Center (FL)
Hoag Memorial Hospital Presbyterian (CA)
Holmes Regional Medical Center (FL)
Holzer Medical Center (OH)
Hopital du Sacre-Coeur de Montreal (Montreal, Quebec, Canada)
Hôpital Maisonneuve – Rosemont (Montreal, Canada)
Hospital for Sick Children (Toronto, ON, Canada)
Hospital Sousa Martins (Portugal)
Hotel Dieu Hospital (Windsor, ON, Canada)
Houston Medical Center (GA)
Huddinge University Hospital (Sweden)
Hurley Medical Center (MI)
Indiana University
Innova Fairfax Hospital (VA)
Institute of Medical and Veterinary Science (Australia)
International Health Management Associates, Inc. (IL)
Jackson Memorial Hospital (FL)
Jersey Shore Medical Center (NJ)
John C. Lincoln Hospital (AZ)
John F. Kennedy Medical Center (NJ)
John Peter Smith Hospital (TX)
Kadlec Medical Center (WA)
Kaiser Permanente Medical Care (CA)
Kaiser Permanente (MD)
Kantonsspital (Switzerland)
Keller Army Community Hospital (NY)
Kenora-Rainy River Regional Laboratory Program (Ontario, Canada)
Kern Medical Center (CA)
Kimball Medical Center (NJ)
King Faisal Specialist Hospital (Saudi Arabia)

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

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

Contents

Contents (Continued)

Foreword

Before using a new method or instrument for *in vitro* diagnostic use, the laboratory must make a preliminary decision about its acceptability. This initial performance check is neither a rigorous characterization of long-term performance nor an evaluation of the many factors that can affect results produced by the device. Rather, this experiment is a quick check to rule out major problems and a starting point for accumulating data and experience that will enable the user to make a final decision. The primary purpose of this document is to help detect performance problems that would warrant immediate correction, referral to the manufacturer, or expanded investigation before a new device is placed into service.

Key Words

Carry-over, comparison of methods, drift, evaluation protocol, experimental design, precision, linearity, linear regression, multiple regression, outlier.

A Note on Terminology

NCCLS, as a global leader in standardization, is committed to achieving global harmonization wherever possible. Harmonization is a process of recognizing, understanding, and explaining differences while taking steps to achieve worldwide uniformity. NCCLS recognizes that medical conventions in the global metrological community have evolved differently in the United States, Europe, and elsewhere; that these differences are reflected in NCCLS, ISO, and CEN documents; and that legally required use of terms, regional usage, and different consensus timelines are all obstacles to harmonization. In light of this, NCCLS recognizes that harmonization of terms facilitates the global application of standards and is an area of immediate attention. Implementation of this policy must be an evolutionary and educational process that begins with new projects and revisions of existing documents.

In order to align the usage of terminology in this document with that of ISO, the term *"Sample"* has replaced the term *"Specimen,"* and is defined as "one or more parts taken from a system, and intended to provide information on the system, often to serve as a basis for decision on the system or its production." The term *"Measurement procedure"* has replaced the term *"Analytical method"* for a set of operations, used in the performance of particular measurements according to a given method; *"Trueness"* has replaced *"Accuracy"* when referring to the closeness of the agreement between the average value from a large series of measurements and to a true value of a measurand. The term "*Accuracy*," to be consistent with ISO terminology, includes both systematic and random components of a (single) measurement. "*Measurement error"//(Error of measurement)* is used instead of *"Total error"* to describe the result of a measurement minus a true value of the measurand.

The term *"Precision,"* is defined as "closeness of agreement between independent test/measurement results obtained under stipulated conditions." As such, it cannot have a numerical value, but may be determined qualitatively as high, medium, or low. For its numerical expression, the term *"Imprecision"* is used, which is the "dispersion of results of measurements obtained under specified conditions." In addition, different components of precision (or imprecision) are defined in EP10, primarily *"Total imprecision,*" *"Within-run imprecision,"* and *"Between-day imprecision."* Other components of measurement error are also described, as determined by the source (nonlinearity, linear drift, and carry-over).

The term *"Within-run imprecision*" is identical to the ISO term *"Repeatability,"* i.e., the closeness of the agreement between results of successive measurements of the same measurand carried out under the same conditions of measurement. The ISO term *"Reproducibility"* describes the closeness of agreement of results of measurements under changed conditions. In this document, reproducibility concepts may include terms such as *"Run-to-run imprecision," "Day-to-day imprecision,"* and *"Lab-to-lab*

imprecision," etc. Reproducibility conditions need to be specified. Measures of nonlinearity, carry-over, and drift are different factors affecting the accuracy of a measurement procedure, and are estimated by this protocol.

At this time, the area committee has chosen not to replace "*Analyte"* with "*Measurand"* (i.e., particular quantity subject to measurement) due to user nonfamiliarity and for the sake of the practicability of the guideline.

The users of EP10 should understand that the fundamental meanings of the terms are similar, and to facilitate understanding, where appropriate, the terms are defined along with their ISO counterpart in the guideline's Definitions section (Section 3).

The Quality System Approach

NCCLS subscribes to a quality system approach in the development of standards and guidelines, which facilitates project management; defines a document structure via a template; and provides a process to identify needed documents through a gap analysis. The approach is based on the model presented in the most current edition of NCCLS HS1- *A Quality System Model for Health Care.* The quality system approach applies a core set of "quality system essentials (QSEs)," basic to any organization, to all operations in any healthcare service's path of workflow. The QSEs provide the framework for delivery of any type of product or service, serving as a manager's guide. The quality system essentials (QSEs) are:

QSEs

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

EP10-A2 addresses the following quality system essentials (QSEs):

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

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

Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition

1 Introduction

This document describes a procedure for the preliminary evaluation of linearity, proportional and constant bias, linear drift, sample carry-over, and precision of a clinical laboratory method. Preliminary evaluations should be performed before new methods are used to test patients' samples and when any modifications of methods are made. This guideline is based on a protocol and analysis method developed for the Technicon SMA analyzer.[1] The experiment is intended primarily for evaluating automated instruments but may be appropriate for kits, manual procedures, or other in vitro diagnostic devices. By repeating a sequence of only ten samples, performance characteristics may be evaluated by plotting the data and performing some simple calculations. Using a statistical technique called multiple linear regression analysis, further information about the factors influencing accuracy (such as sample carry-over and linear drift) can be obtained. Instructions are given for simple data analysis, in case a computer is not available.

The experiment is intended to provide preliminary estimates of those performance characteristics that may be used to determine the ultimate acceptability of the device. The results should be used only to determine whether the device has grossly unacceptable performance.

The following sections outline the materials and procedures to be used. Many variations on this basic experiment are possible (such as extending the number of days or eliminating the priming samples when appropriate). Variations should be dictated by the complexities of the device, the particular characteristics of the method, and the resources available to the user.

2 Scope

Before using a new method, kit, or instrument for in vitro diagnostic use, it is often necessary to make a preliminary decision about its acceptability. This initial performance check is neither a rigorous investigation into the method's long-term performance nor an evaluation of the many factors that can affect results produced by the device. The primary purpose of this document is to help detect problems that are severe enough to warrant immediate correction, referral to the manufacturer, or expanded investigation.

3 Definitions*

Acceptability – Based on individual criteria that sets the minimum operational characteristics for a particular method.

Accuracy – Closeness of agreement between a test result or measurement result and the true value; **NOTE:** In practice, the accepted reference value is substituted for the true value (ISO 3534-1:[3.11]) See **Measurement error**.

Adjusted variance – A statistical manipulation that adjusts the measured variance by subtracting components from other sources of variance. **NOTES**: a) For example, between-run variance is adjusted by subtracting the contribution from within-run variance; b) Appendix C of this document describes a method for determining adjusted variance.

* 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.

Analyte – Component indicated in the name of a measurable quantity; **NOTE**: For example, in the type of quantity "catalytic concentration of lactate dehydrogenase isoenzyme 1 in plasma," "lactate dehydrogenase isoenzyme 1" is the analyte. "Catalytic concentration of" designates the measurand [ISO/FDIS 18153 (January 2002)]. See **Measurand**.

Bias – Difference between the expectation of a test result or measurement result and a true value; **NOTE:** In practice, the accepted reference value is substituted for the true value (ISO 3534-1 [3.13]).

Calibration – The set of operations that establish, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system, or values represented by a material measure or a reference material, and the corresponding values realized by standards; **NOTE:** The term "standard" here refers to measurement standard, not a written standard.

Carry-over – The discrete amount of analyte carried by the measuring system from one sample reaction into subsequent sample reactions, thereby erroneously affecting the apparent amounts in subsequent samples.

Coefficients – In the context of quantitative clinical laboratory methods, the calculated values for the *B* (subscript 0 through 4) values for the linear regression equation; (See **Linear regression**).

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

Comparison of methods – A statistical procedure that is based on data gathered from the paired analysis of the same samples by two different methods. Ideally, one of the methods is a well-accepted or reference method, sometimes called a "gold standard."

Device – An instrument that gives analytical answers as a result of electrical or mechanical measurements on an element, compound, solution, etc.; **NOTE:** The measurement is often made before and after a chemical or physical reaction; the resultant measurement can be manipulated to give a final analytical result.

Difference Plot – A plot of the difference between a measured value and a reference concentration plotted on the y-axis versus the reference concentration on the X-axis; **NOTE**: Often, a dashed line is drawn at zero difference).

Grand mean – Overall mean calculated after multiple runs or days of analysis.

Imprecision – Dispersion of independent results of measurements obtained under specified conditions; **NOTE:** It is expressed numerically as **Standard deviation** or **Coefficient of variation**.

Linear drift – A change in measurement value over time due to factors other than the concentration of the analyte being measured.

Linear regression – A statistical calculation that results in parameters that describe the assumed linear relationship between values of an independent and a dependent variable wherein the independent variable is known exactly; **NOTE:** a) The calculation is based on the mathematical definition of a line ($y = mx + b$); and the mathematical minimization of the vertical distance between each data point and the regression line.

Linearity – The ability (within a given range) to provide results that are directly proportional to the concentration {amount} of the analyte in the test sample*;* **NOTE:** Linearity typically refers to overall system response (i.e., the final analytical answer rather than the raw instrument output.

Matrix – All components of a material system, except the analyte.

Measurand – particular quantity subject to measurement [VIM:1993, 2.6]. See **Analyte**.

Measurements procedure – A set of operations, described specifically, used in the performance of particular measurements according to a given method; **NOTE:** Formerly, in this document, the term **Analytical method** was used.

Measurement error//Error of measurement – The result of a measurement minus a true value (or accepted reference value) of the measurand; **NOTE:** a) Formerly, the term **Total error** was used.

Multiple linear regression – A statistical calculation that provides multiple parameters to different factors. See linear regression (above) for more information.

Outlier – The observation in a sample, so far separated in value from the remainder as to suggest that is may be from a different population, or the result of an error in measurement. (ISO 3534-1 [2.64])

Pooled within-run variance – The overall within-run variance. In this procedure, because all the days have the same number of data points, the daily within-run variances are averaged.

Precision – The closeness of agreement between independent test results obtained under stipulated conditions (ISO 3534-1 [3.14]); **NOTE:** Precision is not represented as a numerical value but is expressed quantitatively in terms of imprecision–the SD or the CV of the results in a set of replicate measurements. See **Imprecision**.

Regression multiplier – Factors derived from the experimental design and used (in Data Summary Sheet #4) to multiply the observed data.

Repeatability//(Within-run precision) – Closeness of the agreement between results of successive measurements of the same measurand carried out under the same conditions of measurement.

Reproducibility//(Total imprecision) – imprecision under reproducibility conditions; **NOTE**: "Total imprecision" refers to the experimental conditions outlined in this document; See **Reproducibility conditions.**

Reproducibility conditions – Conditions where test results are obtained with the same method on identical test items under different settings, e.g. in different laboratories, with different operators, using different equipment; **NOTE**: In this document, the reproducibility conditions are defined in the experimental model in Sections 6 and 7.

Sample – One or more parts taken from a system, and intended to provide information on the system, often to serve as a basis for decision on the system or its production; [ISO/DIS 15190], [prEN ISO/DIS 15189.2].

Scale factor – The difference between the label value of sample 1 (low concentration) and 2 (mid-concentration) of the samples used in this protocol.

Standard deviation, SD, σ – 1) A measure of variability/dispersion that is the positive square root of the population variance; **NOTE**: A number of formulae exist for calculating standard deviation; users of this protocol should use that which is shown in Appendix C.

Total imprecision – See **Reproducibility**.

Trueness – The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value (ISO 3534-1 [3.12]); **NOTE:** It is expressed numerically as **Bias**.

t-test – An abbreviation for the Student's t-test, a statistical test based on the Student's t-distribution; **NOTE:** The Student's t-test is most commonly used in evaluating whether two sample means are different.

Variance, σ^2 – 1) A measure of dispersion of observations which is the sum of the squared deviations of observations from their average divided one less than the number of observations. (ISO 3534-1 [2.33])

$$\sigma^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

4 Symbols

The following symbols are used in the calculations (several are not standard, but they are used for the sake of convenience):

•—multiplication function.

x—expected or labeled value of the samples used in this protocol. It is plotted on the horizontal axis of the difference plot.

y-x—observed or experimentally obtained difference for the samples used in this protocol. It is plotted on the vertical axis of the difference plot.

C—assigned value for each concentration.

D— the difference between the observed value (*y*) (or the average of a series of values) and the assigned value (*C*).

N—number of data points at any of the three levels used in this protocol.

t—calculated statistic to test the significance of the bias; commonly referred to as Student's *t*-test.

R—pooled within-run variance. For more information, see pooled within-run variance in the Definitions section.

S—variance of daily means. It is calculated as the squared standard deviation of the daily means.

$\boldsymbol{SD_L}$—total standard deviation for any level. It is calculated using equation 1 in Appendix C.

$\mathbf{s_{y.x}}$—This residual standard deviation term, also called the root mean square from analysis of variance is an estimate of the standard deviation of the test method with all effects of parameters that have been modeled removed.

T—adjusted between-day variance. It is calculated as the variance of daily means minus one-third of the pooled within-run variance.

U—total i variance, which is the pooled within-run variance plus the corrected between-day variance.

V—total standard deviation, calculated using the standard deviation formula in Appendix A.

W—total *CV*% (see Definitions section) specified within-run based on the calculation in Appendix A.

5 Materials

Three stable pools of analyte that span the range claimed or the medically relevant range (not to exceed the claimed range) for the test method are required. Such materials may be obtained commercially (e.g., control materials) or may be made from patient sample pools. The concentration of the midlevel pool must be exactly halfway between the high- and low-level concentrations. An efficient way to create the midlevel is by mixing equal parts of the high- and low-level pools. To prime the device in each run, more of the midlevel material is needed than the other two levels.

The sample matrix must be compatible with the requirement of the method. Interferences may be seen if an inappropriate matrix is used. The manufacturer's guidelines for appropriate sample matrices should be followed.

If the high and low pools do not adequately span the useful range, the high-level pool may be spiked when appropriate. Similarly, the low-level pool may be diluted. In both cases, consideration must be given to potential matrix effects.

Choose the high and low pools carefully for their stability over the course of the experiment. If the analyte is stable, prepare a sufficient amount of all three materials to last for the entire evaluation. If the material is unstable, use frozen aliquots or control material reconstituted daily.

6 Calibration and Sequence of Samples in a Run

This experiment consists of a series of analytical runs made over several days. A “run” is a sequence of samples analyzed consecutively without interruption, unless the recommended operation requires such interruption. In particular, there should not be any recalibration within this recommended sequence. Otherwise, calibration should be performed according to the manufacturer’s recommendations. The first sample is used to prime the system, that is, to create a consistent carry-over on the third sample in the sequence. Sample carry-over is assessed by mathematical analysis. The results, starting with the second sample, are used for data analysis from each run. The following specific sequence must be analyzed in a run without change, interruption, or intervening samples: Mid, Hi, Low, Mid, Mid, Low, Low, Hi, Hi, Mid. If any of the last nine samples is rejected, lost, or not reported for whatever reason, the entire run must be repeated. While it is necessary to discard all the data from such an incomplete run, every effort must be made to determine the reason for the rejection. This sequence was specifically designed to allow the nearly uncorrelated estimation of the effects of nonlinearity, sample carry-over, proportional and constant bias, and linear drift.

Note that the multiple regression procedure in this guideline is designed to analyze for *sample* carry-over. Many modern analyzers are random access analyzers, i.e., required batch mode of the above sequence may not be carried out. This invalidates the assumptions used in the multiple regression and may give invalid results. The protocol and multiple regression procedure can be adjusted to estimate both reagent carry-over and sample carry-over but is beyond the scope of this document.[2]

7 Number of Days and Runs

At least one run per day must be performed for at least five days. More runs and days will increase confidence in the results. To incorporate data from additional days or runs, the laboratorian should use Appendix C. To simplify the data analysis, the same number of runs should be performed each day. The examples in this document use one run a day over five days, the minimum required for the experiment.

8 Preliminary Procedures

Before beginning the experiment, all laboratorians operating the device must become familiar with its operation. Practicing start-up, calibration, error recovery, cleaning, shutdown, reagent preparation, and all other operational factors is essential before beginning the acceptability check. Where appropriate, manufacturer's instrument training and installation should be completed before starting this protocol.

9 Collection and Recording of Data

Appendix A contains example data recording sheets. They can be used as a visual aid to ensure that the samples were run in their proper sequence and to record the means of the observations at each of the three levels. Using these or appropriate variations will greatly facilitate the data analysis. On Data Summary Sheet #1, the full ten-sample sequence is recorded. The first sample is a prime and only the nine subsequent sample results are analyzed.

10 Initial Data Plotting and Inspection

The data from this experiment may be analyzed in a variety of ways. However, the data must first be plotted to get a picture of the spread of the observations within each concentration level and to view the results simultaneously at all three levels. The graph can convey an impression of the, precision, bias, and nonlinearity. Visually assessing the data is necessary because there is little "power" in any statistical summary from such a small set of data.

10.1 Difference Plot of Data Versus Concentration

Figure 1 shows the plotting that should be used for the data. The *X*-axis is the expected (labeled) concentration. The *Y*-axis represents the observed difference of the assay from reference. A zero difference line on the plot should be drawn for reference. Each individual point (15 at each level) should be plotted.

10.2 Visual Inspection for Outliers

Careful examination of the data plot for outliers (i.e., single points detached from the main cluster of points at a concentration level) involves individual judgment. When there is difficulty in deciding whether a point is an outlier, it should be left in the data analysis. The treatment of outliers is particularly critical in such a small experiment because a detached point can greatly affect the analysis of the data.

If an outlier is found, every effort should be made to determine the cause; it may indicate a fundamental problem. If it is observed in more than one run, the appropriate actions are to: (1) troubleshoot, (2) include the apparently offending points in the subsequent analysis, or (3) terminate this preliminary investigation and begin an expanded evaluation of error sources. If the outliers are included in the data analysis, the conclusions derived from the analysis may be due largely and disproportionately to the outliers. A single run that has an outlier may be replaced with another run.

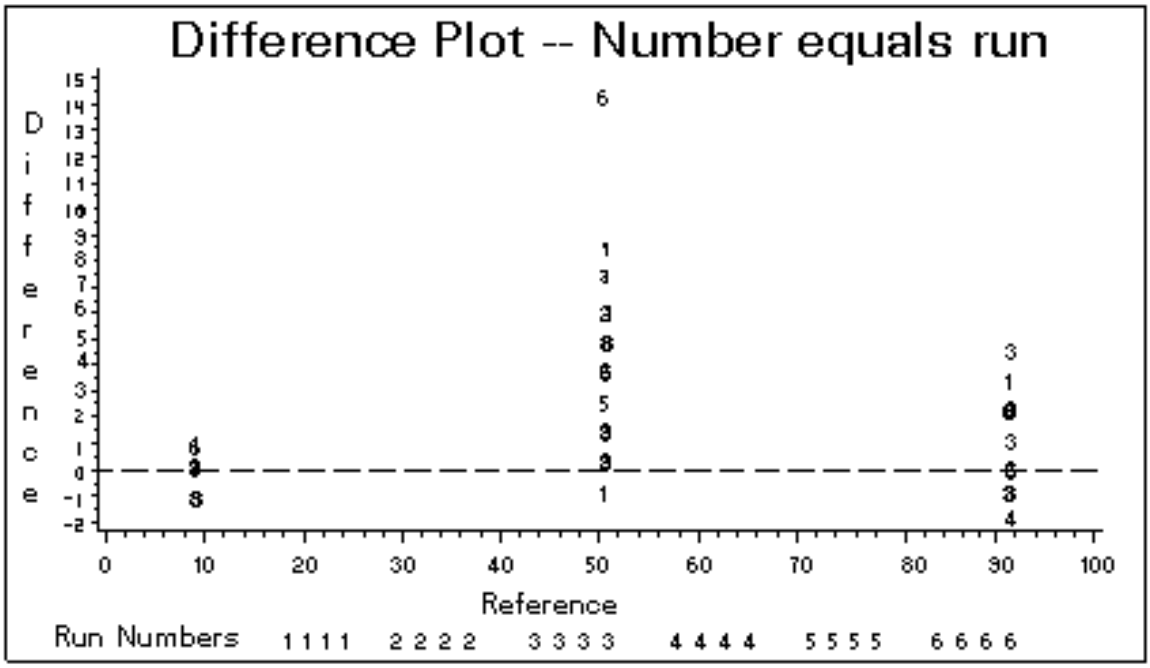

Figure 1. **Raw Data Difference Plot BUN Data.** (One run on each of six days; [Day 6 declared an outlier day].)

10.3 Visual Inspection for Linearity

The difference plot should be examined for any indication of nonlinearity. As an example, Figure 1 exhibits nonlinearity because the midlevel results are generally higher than either the low or high level results. If the data still do not appear linear, then it is best to evaluate whether the multiple linear regression analysis coefficient (B_3) (included in each of the worksheets) is significant on more than one run or in the summary analysis of all runs and then expand the investigation to include a full evaluation of linearity (such as that described in the current version of NCCLS document EP6—*Evaluation of the Linearity of Quantitative Analytical Methods*). Again, only considerable nonlinearity can be detected with this preliminary experiment because of the small amount of data collected. The degree of nonlinearity that may be acceptable is left to the judgment of the individual laboratory.

11 Analysis of the Data for Imprecision

The total imprecision (standard deviation) can be estimated in one of two ways. The first and simplest is to compute the standard deviation (SD) of all observations within each concentration. This will produce a "simple" total standard deviation estimate for each level and may be used as the initial statistic for the preliminary examination of imprecision of the test method.

The "simple" estimators of the total standard deviations (i.e., one for each level) will often slightly underestimate the total standard deviations as calculated below. A formula to calculate the difference between the two calculation methods is available.[3]

The "standard" estimator of the total standard deviation may be computed by estimating the components of variance due to within-run, between-run (if more than one run is done per day), and between-day factors, as given by the formulas and procedures described in Appendix C. These components of variance should then be added and the square root calculated to yield the "corrected" total standard deviation. The relative sizes of the components may then be examined to investigate the sources of imprecision.

Further discussion of the use of the individual variance components is beyond the scope of this document. If it is necessary to go into this type of expanded total imprecision breakdown, and examination of the individual components does not yield information leading to the source of imprecision, then the data analysis should be halted. The indicated courses of action include contacting the manufacturer or performing an extended 10- to 20-day precision evaluation, such as that described in the current edition of NCCLS EP5—*Precision Performance of Clinical Chemistry Devices*.

11.1 Interpretation

For each of the three levels, one needs to provide imprecision goals. One can then compare the estimated total imprecision to its goal for each level. The actual comparison is to reject values higher than the goal and to accept values lower than the goal. (See Data Sheet #3 for an example.) Alternatively, one could perform a statistical test. One should realize that the estimated total imprecision is a point estimate. This means that if one were to repeat the experiment many times, a range of values would be produced. However, point estimates occurring close to the real value would occur more often than point estimates far away from the real value. While the construction of imprecision goals is beyond the scope of this document, one should also note that imprecision goals can be stated either as point estimates or maximum expected imprecision (See the current version of NCCLS document EP11—*Uniformity of Claims for In Vitro Diagnostic Tests*). To help define imprecision goals, one could use the performance of existing or similar analytes or manufacturer's labeling claims.

12 Preliminary Assessment of Bias

Bias can be estimated by the difference between the observed mean values and the assigned values at each concentration.

12.1 Assigned Values

The estimates of bias will be only as good as the degree to which the pools emulate the samples being tested and the trueness of the values assigned to the pools. The proper method to assign these values depends on the nature of the pools used.

12.1.1 Pooled Patient Samples

Analysis of an aliquot should be done on each day of the evaluation by another method known to be accurate (a reference method is ideal). The mean should be calculated to determine the assigned value.

12.1.2 Control Samples

These fluids may not behave the same as patient samples and are therefore not valid indicators of method bias. However, they can be used to verify that the results are consistent with the manufacturer's expectation for the method.

The value assigned by the manufacturer, if available, or the peer group mean for the system being tested should be used. A value cannot be assigned by comparison with another method because of potential matrix effects that may affect either method.

12.2 Calculation of Bias

12.2.1 Control or Patient Sample Pools

The observed mean should be calculated from the results recorded on Data Summary Sheet #2. For each of the three concentrations, the assigned value is subtracted from the observed mean to obtain an estimate of bias:

Bias = Observed Mean - Assigned Value.

12.2.2 Individual Patient Samples

The results should be recorded on Data Summary Sheet #7, grouped according to the sample concentration (high, medium, low). The difference between the two methods should be calculated for each sample and the average difference (bias) for each group determined, as indicated on the worksheet.

12.3 Interpretation

The observed difference could be due to one or more factors, including method trueness, incorrect standardization, measurement uncertainty, interferences, matrix effect, drift, sample carry-over, and incorrectly assigned values.

The estimated bias should be compared with your allowable goal for bias.

13 Full Data Analysis Procedures

The design of the experiment allows estimation of the effect of the slope (proportional bias), intercept (constant bias), sample carry-over, nonlinearity, and linear drift. Section 11 and Appendix C can be used to estimate the components of variance. Appendix C also includes a multiple regression procedure that may be used if the initial data analysis indicates possible unacceptable performance of the method. The multiple regression procedure consists of the formulae in Appendix C and Data Summary Sheets #4, #5, and #6. It is, however, necessary to go through this procedure only if use of the details concerning the individual components contributing to measurement error is desired.

13.1 A Comment on the Model

The EP10 multiple regression model might not always seem to be appropriate. For example, sample carry-over may be virtually impossible due to the design of an instrument. One solution to this of course is to contact a statistician to develop a different design. Yet using the NCCLS EP10 design in the above case where sample carry-over is unlikely could still yield valuable information for the following reason. From a model calculation standpoint, it is irrelevant whether effects are likely or unlikely, since all parameters in the model are automatically (i.e., mathematically) estimated. This means that variability in the data is explained as either nonzero parameter effects or as imprecision (residual error). So it is possible that some effect other than sample carry-over, which one has not thought of, is estimated as sample carry-over by the model. This would be a signal to investigate the measurement procedure to determine the origin of this error source.[4]

13.2 Summarizing the Five Runs

The *t*-test procedure that had been used to summarize the five runs has been changed. With only five runs, the problem is that variation in a parameter estimate from run to run could be due to sampling variation (implies that the parameter is a constant that cannot be perfectly estimated with only ten samples) or that the parameter is truly changing from run to run. Of the possible ways to summarize the runs, the

subcommittee chose a one-sample sign test.[5] This is likely to be significant in many cases as it requires all estimates values to have the same sign for the five runs (1 is subtracted from the slope estimates).

If a result has been detected as statistically significant, it may still be of no practical importance as an error source. Users must decide on how much error to allow to each error source.

14 An Alternative Procedure

A modification of the protocol can be carried out that allows the estimation of three interfering substances in addition to all of the other parameters estimated.[6]

15 Summary

The committee believes this procedure provides the maximum amount of information on performance that is available from a minimum expenditure of time and material. To accomplish this, the data must be collected with utmost care so that outliers do not destroy the effectiveness of the experiment. Also, the data analysis procedures are designed so that each data point serves several purposes and all possible information is extracted from the selected sequence of concentration values. Detailed imprecision assessment requires the use of the fully nested analysis of variance (ANOVA) with the properly corrected variance components estimated (Worksheet #3). The assessment of nonlinearity, linear drift, sample carry-over, and bias requires the use of multiple linear regression, simplified to arithmetic calculations appropriate only to such a specifically designed experiment (Data Summary Sheets #4, #5, and #6).

The final decision as to the acceptability of the method should be based on the medical usefulness of the assay results and currently accepted standards of laboratory practice. Although limited by a small amount of data, this protocol can be a powerful tool that allows evaluation of the minimum acceptability of an instrument or method.

References

[1] Daniel C. Calibration designs for machines with carry-over and drift. *J Quality Technol*. 1975;7:103-108.

[2] Krouwer JS, Stewart WN, and Schlain B. A multifactor experimental design for evaluating random-access analyzers. *Clin. Chem*. 1988;33:1894-1896.

[3] Krouwer J. Observations on comparisons of within-run and day-to-day precision. *Clin Chem*. 1981;27:202.

[4] Krouwer J. Assay Development and Evaluation. *A Manufacturer's Perspective*. AACC Press; Washington, D.C.; 2002.

[5] Dixon WJ, Massey FJ. *Introduction to Statistical Analysis*. 4th ed. New York: McGraw-Hill; 1983:385-388.

[6] Krouwer J, Monti K. Modification of NCCLS EP10 to include interference screening. *Clin Chem.* 1995;41:325-326.

Appendix A. Preliminary Performance Acceptability Check

Data Summary Sheet #1: Individual Run Raw Data

Method/instrument:
Assigned Values (C):
Low:
Mid:
High:
Run#

Source of pools:
Analyte:
Date/time:
Operator:
Day#:

Write each observed value (*Y*) *twice*, once in each of the appropriate columns:

Sequence Number	Level	Value	Low Values	Mid Values	High Values	Transformed Values
0	Mid			Do not record value		0
1	High					1
2	Low					-1
3	Mid					0
4	Mid					0
5	Low					-1
6	Low					-1
7	High					1
8	High					1
9	Mid					0

Sum			
Mean			
Within-run standard deviation*			
With-run variance†			

* See Appendix C for basis of calculations. Standard deviations use equation 1 of Appendix C or a calculator that calculates the standard deviation with n - 1 as the denominator.
† The variance equals the standard deviation squared.

Appendix A. (Continued)

Data Summary Sheet #2: **Calculations of Bias: All Runs**

Method/instrument:
Assigned Values (C):
Low:
Mid:
High:

Source of pools:
Analyte:
Date/time:
Operator:
Day#:

Summary of observed values for each run from data Summary Sheet #1.

	Low		Mid		High	
Day	**Within-Run Standard Deviation**	**Mean**	**Within-Run Standard Deviation**	**Mean**	**Within-Run Standard Deviation**	**Mean**
1						
2						
3						
4						
5						

Grand mean (*Y*)		
Labeled Value (C)		
Bias = *D* = (*Y* - *C*)		
Your allowable bias		

Daily within-run standard deviations and means from Data Summary Sheet #1.

Appendix A. (Continued)

Data Summary Sheet #3: **Calculation of Imprecision: All Runs**

Method/instrument:
Assigned Values (C):
Low:
Mid:
High:

Source of pools:
Analyte:
Date/time:
Operator:
Day#:

	Low	Mid	High
(*R*) Pooled within-run variance*			
(*S*) Variance of daily means†			
(*T*) Adjusted between-day variance‡ (*S*) - (*R*)/3			
(*U*) Total variance (*R*) + (*T*)			
(*V*) Total standard deviation = $\sqrt{U}$			
Grand mean value (from Worksheet #2)			
(*W*) Total *CV*% = (*V*)/*Grand Mean Value* • 100%			
Your allowable *CV*%			
Accept or reject			

If estimates of imprecision are satisfactory, then stop here. If an evaluation of the components of imprecision is desired, continue to Data Summary Sheets #4, #5, and #6.

* Because all days have the same number of data points, it is permissible to simply take the mean within-run variance for all accepted runs at each level.
† Calculated as the variance of daily means from Data Summary Sheet #2 for each level.
‡ If less than zero, set equal to zero.

Appendix A. (Continued)

Data Summary Sheet #4: Multiple Regression Calculations (One Data Sheet Per Run)

Method/instrument:
Assigned Values (C):
Low:
Mid:
High:
Run#

Source of pools:
Analyte:
Date/time:
Operator:
Day#:

Scale factor: (*Label Value* [*Mid*] - *Label Value* [*Low*])

	Slope				Carry-over		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1		• 139 =				• 26 =	
2		• -96 =				• 130 =	
3		• 11 =				• -102 =	
4		• 8 =				• 8 =	
5		•-117 =				• -4 =	
6		•-126 =				• -126 =	
7		• 100 =				• -126 =	
8		• 100 =				• 100 =	
9		• -19 =				• 94 =	
		Total =				Total =	
		Total/678 =				Total/678 =	
			(slope, *B1*)				(carry-over, *B2*)

	Nonlinearity				Drift		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1		• 87 =				• -52 =	
2		• 96 =				• -34 =	
3		•-237 =				• -22 =	
4		•-234 =				• -16 =	
5		• 117 =				• 8 =	
6		• 126 =				• 26 =	
7		• 126 =				• 26 =	
8		• 126 =				• 26 =	
9		•-207 =				• 38 =	
		Total =				Total =	
		Total/678 =				Total/678 =	
			(nonlinearity, *B3*)				(linear drift/test, *B4*)

The above coefficients are multipliers derived from the experimental design and should not be changed.

Appendix A. (Continued)

Data Summary Sheet #5: *t*-Statistic for Regression Coefficients: **Single Run**

Method/instrument:
Assigned Values (C):
Low:
Mid:
High:
Run#

Source of pools:
Analyte:
Date/time:
Operator:
Day#:

(1) Compute intercept: B_0 = _____ is the average of the y values from each run.
(2) Compute standard error or estimate: Use multiple linear regression equation below to calculate Y_j^* (predicted value of Y)

Assay #	Assay Transformed Value	Observed Value	Computed *Y*	Residual	Residual Squared
j	x_j	y_j	Y_j^*	$y_j - Y_j^*$	$(y_j - Y_j^*)^2$
0					
1					
2					
3					
6					
9					
				$S(y_j - Y_j^*)^2$=	
				$S_{y \cdot x}$ =	

$$S_{y \cdot x} = \sqrt{\frac{\Sigma\left(y_j - Y_j^*\right)^2}{4}}$$

$Y_j^* = B_0 + B_1 \bullet x_j + B_2 \bullet x_{j-1} + B_3(x_j^2 - 2/3) + B_4 \bullet t$ (see Appendix C) where x_j is the transformed value for x (i.e., low = -1, mid = 0, and high = +1) and the t multiplied by B_4 is time (-4 through +4).

(3) t-Statistics

(A) Regression Parameters	(B) Regression Equation Value	(C) Standard Error Value	(D) Standard Error	(E) *t*-Statistic *t = B/D*
B_1		$S_{y \cdot x} \bullet 0.4135$ =		
B_2		$S_{y \cdot x} \bullet 0.4135$ =		
B_3		$S_{y \cdot x} \bullet 0.7099$ =		
B_4		$S_{y \cdot x} \bullet 0.1330$ =		

t for 4 degrees of freedom is significant ($p < 0.01$) if $t > 4.6$ or $t < -4.6$.

Appendix A. (Continued)

Data Summary Sheet #6: **Multiple Regression Summary: All Runs**

Method/instrument:
Assigned Values (C):
Low:
Mid:
High:

Source of pools:
Analyte:
Date/time:
Operator:
Day#:

Day	Intercept (B_0adj)	Slope (B_1adj)	%Carry-over (B_2adj)	Nonlinearity (B_3)	Drift (B_4)	$S_{y \cdot x}$
1 Value						
t						
2 Value						
t						
3 Value						
t						
4 Value						
t						
5 Value						
t						

t-value comes from column E in Data Summary Sheet #5.

t is significant ($p < 0.01$) if $t > 4.6$ or $t < -4.6$.

Appendix A. (Continued)

Data Summary Sheet #7: **Individual Patient Sample Comparisons (Optional)**

Device X (old): Analyte:
Device Y (new):

	Low Concentration			Medium Concentration			High Concentration		
Day	***Y***	***X***	**(Bias) *Y-X***	***Y***	***X***	**(Bias) *Y-X***	***Y***	***X***	**(Bias) *Y-X***
1									
2									
3									
4									
5									
Average Concentration of X =									
Average Bias =									
% Bias = (Average Bias/Average Concentration) 100 =									

Comments:

Appendix B. Example Use of Data Sheets

General Instructions

(Note: Calculated values found in the examples may differ slightly due to rounding variations.)

The example is data from blood urea nitrogen (BUN) assays on a small analyzer. Six actual runs were made. Run Number 5 was rejected because of an outlier BUN value of 65 mg/dL when the mean was 50.5. The sixth run replaced the rejected run. The basis for each of the calculations is found in Appendix C The following is an item-by-item discussion of the data analysis sheets listed by number.

(1) One Data Summary Sheet #1 is required for each of the five runs. The data that are actually used for analysis are data sequence numbers 1 through 9. The demographic information should be entered at the header, which is similar for each of the data summary sheets. Each analysis value should be entered twice, once in the value column and once in the appropriate low, mid, or high column. Note that each of the values is coded or transformed to -1 (low), 0 (mid), and +1 (high) for statistical analysis by the linear regression equation in Data Summary Sheet #5. These transformed values can also be used for plotting the values on a scatter plot (for the *y* axis, the assay values should be used; for the *x* axis, the assigned values or the transformed values should be used). The mean and standard deviation calculations are simple sums of the columns. The mean = sum/number of data points, and the standard deviation is calculated by a calculator using *n* - 1 degrees of freedom (number of data points - 1) in the denominator. The variance equals the standard deviation squared. The sixth Data Summary Sheet #1 is an example of a rejected run that had an outlier.

(2) Data Summary Sheet #2 is used to summarize the within-run standard deviations (in concentration units) and the means for each level and each run. These data are then used to calculate the grand mean and bias for each level. The adequacy of the demonstrated bias should be evaluated in comparison to allowable bias, keeping in mind the possible effects of the sample matrix on the analysis.

(3) Data Summary Sheet #3 is used to calculate the total imprecision expressed as standard deviation and coefficient of variation (CV%). These values should then be compared to the allowable imprecision and a decision to accept or reject should be made. The (*R*) value, or pooled within-run variance, is calculated by a simple variance calculation or as in the first footnote (*) by saying that this can be a simple average of each run's within-run variance because each run has the same number of data points. The (*S*) value is calculated using a variance formula (Formula 1 of Appendix C, without taking the square root). The (*V*) value is the square root of the (*U*) value.

If the bias, imprecision, and plot of the data appear to be within specifications, the instrument or method can be judged as preliminarily acceptable, and this initial screening of the quality of the instrument/method is finished. Further and ongoing evaluation of the usefulness of the instrument/method in the individual laboratory is based on other, more long-term evaluations. If specific or general problems with the instrument/method are found, then the manufacturer should be contacted and given a description of the outcome of the preliminary evaluation. More complex evaluations may also be performed using one of the other NCCLS evaluation protocols. (See the Related Publications section). To evaluate the statistical components of the variances, Data Summary Sheets #4, #5, and #6 should be used. Decisions on the significance of the error terms from the multiple linear regression calculations are made from the data summary found on Data Summary Sheet #6.

(4) Data Summary Sheet #4 is used to calculate the individual components of the measurement error. Components are calculated for intercept (B_0), slope (B_1), sample carry-over (B_2), nonlinearity (B_3), and linear drift (B_4). This technique utilizes multiple linear regression with five terms (the four terms

just mentioned plus an intercept term). Data Summary Sheet #5 will introduce the fifth, or intercept, term. The calculation of the error terms depends on the exact sequence of samples in the analysis scheme and the multiplication by coefficients that are precalculated and listed in Data Summary Sheet #4. The calculation terms B_1 through B_4 are used in Data Summary Sheet #5. There is one Data Summary Sheet #4 for each run.

(5) Coefficients derived from Reference 2 were used in the design of this experiment. Data Summary Sheet #5 is the calculation sheet for Student's *t*-statistics, which compares the regression parameter values with the standard error for that run. The resultant *t*-statistic is then compared with significant values of *t* with 4 degrees of freedom. If the absolute value of the calculated value of *t* exceeds 4.6 ($t > 4.6$ or $t < -4.6$), then the differences are significant for 99 out of 100 times ($p < 0.01$). The intercept, based on the transformed *x* data, is simply the mean of all the data. The $S_{y \cdot x}$ is calculated by computing the *Y* value using the multiple linear regression equation, where *x* is the transformed value of *x* (low -1, mid 0, high +1), B_0 is the calculated intercept, and B_1, B_2, B_3, and B_4 are the estimated regression parameters calculated in Data Summary Sheet #4. The last parameter calculated in $B_4 \bullet t$, here, *t* is time factor, whose value varies with each assay number (i.e., $t = -4, -3, -2, -1, 0, 1, 2, 3,$ and 4). The difference in the calculated *Y* value and the individual paired *y* assay value is squared, summed, and the $S_{y \cdot x}$ calculated using the accompanying formula. The calculated $S_{y \cdot x}$ is used to calculate the standard error value shown in the (D) column. There is one Data Summary Sheet #5 for each run.

(6) Data Summary Sheet #6 is a summary of the data on Data Summary Sheets #4 and #5. Decisions on the significance of the various components (slope, carry-over, nonlinearity, or drift) are made using the *t* values compared to the significance levels you desire (e.g., at $p = 0.01$, *t* must be < 4.6 and > -4.6).

Appendix B. (Continued)

Data Summary Sheet #1a: Individual Run Raw Data

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 8 August 1988
Operator: RBP
Day#: 1

Write each observed value (*Y*) *twice*, once in each of the appropriate columns:

Sequence Number	Level	Value	Low Values	Mid Values	High Values	Transformed Values
0	Mid	51		Do not record value		0
1	High	92			92	1
2	Low	9	9			-1
3	Mid	54		54		0
4	Mid	56		56		0
5	Low	10	10			-1
6	Low	9	9			-1
7	High	92			92	1
8	High	95			95	1
9	Mid	59		59		0

Sum	28	169	279
Mean	9.3	56.3	93.0
Within-run standard deviation†	0.58	2.52	1.73
Within-run variance‡	0.33	6.33	3.00

* BUN, blood urea nitrogen
† See Appendix C for basis of calculations. Standard deviations use equation 1 of Appendix C or a calculator that calculates the standard deviation with *n* - 1 as the denominator.
‡ The variance equals the standard deviation squared.

Appendix B. (Continued)

Data Summary Sheet #1b: Individual Run Raw Data

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 9 August 1988
Operator: RBP
Day#: 2

Write each observed value (*Y*) *twice*, once in each of the appropriate columns:

Sequence Number	Level	Value	Low Values	Mid Values	High Values	Transformed Values
0	Mid	51		Do not record value		0
1	High	92			92	1
2	Low	9	9			-1
3	Mid	54		54		0
4	Mid	54		54		0
5	Low	9	9			-1
6	Low	8	8			-1
7	High	12			91	1
8	High	95			92	1
9	Mid	56		56		0

Sum	26	164	275
Mean	8.7	54.7	91.7
Within-run standard deviation†	0.58	1.15	0.58
Within-run variance‡	0.33	1.33	0.33

* BUN, blood urea nitrogen
† See Appendix C for basis of calculations. Standard deviations use equation 1 of Appendix C or a calculator that calculates the standard deviation with *n* - 1 as the denominator.
‡ The variance equals the standard deviation squared.

Appendix B. (Continued)

Data Summary Sheet #1c: **Individual Run Raw Data**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 10 August 1988
Operator: RBP
Day#: 3

Write each observed value (*Y*) *twice*, once in each of the appropriate columns:

Sequence Number	Level	Value	Low Values	Mid Values	High Values	Transformed Values
0	Mid	51		Do not record value		0
1	High	93			93	1
2	Low	9	9			-1
3	Mid	54		54		0
4	Mid	54		54		0
5	Low	9	9			-1
6	Low	9	9			-1
7	High	92			92	1
8	High	96			96	1
9	Mid	58		58		0

Sum	27	16	281
Mean	9.0	55.3	93.7
Within-run standard deviation†	0.00	2.31	2.08
Within-run variance‡	0.00	5.34	4.33

* BUN, blood urea nitrogen
† See Appendix C for basis of calculations. Standard deviations use equation 1 of Appendix C or a calculator that calculates the standard deviation with *n* - 1 as the denominator.
‡ The variance equals the standard deviation squared.

Appendix B. (Continued)

Data Summary Sheet #1d: **Individual Run Raw Data**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 11 August 1988
Operator: RBP
Day#: 4

Write each observed value (*Y*) *twice*, once in each of the appropriate columns:

Sequence Number	Level	Value	Low Values	Mid Values	High Values	Transformed Values
0	Mid	51		Do not record value		0
1	High	90			90	1
2	Low	9	9			-1
3	Mid	54		54		0
4	Mid	55		55		0
5	Low	9	9			-1
6	Low	9	9			-1
7	High	92			92	1
8	High	94			94	1
9	Mid	52		52		0

Sum	27	161	276
Mean	9.0	53.7	92.0
Within-run standard deviation†	0.00	1.53	2.00
With-run variance‡	0.00	2.34	4.00

* BUN, blood urea nitrogen
† See Appendix C for basis of calculations. Standard deviations use equation 1 of Appendix C or a calculator that calculates the standard deviation with *n* - 1 as the denominator.
‡ The variance equals the standard deviation squared.

Appendix B. (Continued)

Data Summary Sheet #1e: **Individual Run Raw Data**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 15 August 1988
Operator: RBP
Day#: 6

Write each observed value (*Y*) *twice*, once in each of the appropriate columns:

Sequence Number	Level	Value	Low Values	Mid Values	High Values	Transformed Values
0	Mid	52		Do not record value		0
1	High	92			92	1
2	Low	9	9			-1
3	Mid	52		52		0
4	Mid	55		55		0
5	Low	9	9			-1
6	Low	9	9			-1
7	High	92			92	1
8	High	94			94	1
9	Mid	53		53		0

Sum	27	160	278
Mean	9.0	53.3	92.7
Within-run standard deviation†	0.00	1.53	1.15
Within-run variance‡	0.00	2.34	1.32

* BUN, blood urea nitrogen
†See Appendix C for basis of calculations. Standard deviations use equation 1 of Appendix C or a calculator that calculates the standard deviation with *n* - 1 as the denominator.
‡The variance equals the standard deviation squared

Appendix B. (Continued)

Data Summary Sheet #1 (Rejected-Outlier): **Individual Run Raw Data**

Method/instrument: BUN* Analyzer O
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 12 August 1988
Operator: RBP
Day#: 5

Write each observed value (*Y*) *twice*, once in each of the appropriate columns:

Sequence Number	Level	Value	Low Values	Mid Values	High Values	Transformed Values
0	Mid	52		Do not record value		0
1	High	91			91	1
2	Low	8	8			-1
3	Mid	52		52		0
4	Mid	55		55		0
5	Low	10	10			-1
6	Low	9	9			-1
7	High	92			92	1
8	High	94			94	1
9	Mid	65		65		0

Sum	27	172	277
Mean	9.0	57.3	92.3
Within-run standard deviation†	1.00	6.81	1.53
Within-run variance‡	1.00	46.38	2.34

NOTE: This run was deleted based on visual inspection of the difference plot of all data. No reason could be found for the cause of the outlier, which in itself is a cause for concern.

* BUN, blood urea nitrogen
† See Appendix C for basis of calculations. Standard deviations use equation 1 of Appendix C or a calculator that calculates the standard deviation with *n* - 1 as the denominator.
‡ The variance equals the standard deviation squared.

Appendix B. (Continued)

Data Summary Sheet #2: Calculations of Bias: All Runs

Method/instrument: BUN* Analyzer A
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0

Source of pools: Controls
Analyte: BUN
Date/time:
Operator: RBP
Day#: 2

Summary of observed values for each run from data Summary Sheet #1.

	Low		Mid		High	
Day	**Within-Run Standard Deviation**	**Mean**	**Within-Run Standard Deviation**	**Mean**	**Within-Run Standard Deviation**	**Mean**
1	0.58	9.33	2.52	56.33	1.73	93.00
2	0.58	8.67	1.15	54.67	0.58	91.67
3	0.00	9.00	2.31	55.33	2.08	93.67
4	0.00	9.00	1.53	53.67	2.00	92.00
5	0.00	9.00	1.53	53.33	1.15	92.67

Daily within-run standard deviations and means from Data Summary Sheet #1a-e.

Grand mean (*Y*)	9.0	54.7	92.6
Labeled Value (C)	9.0	50.5	92.0
Bias = *D* = (*Y* - *C*)	0.0	4.2	0.6
Your allowable bias	±2	±4	±5

Evaluation: *The allowable bias is set by the laboratory director. The bias for midlevel sample is slightly high, while the low and high levels are very good.*

* BUN, blood urea nitrogen

Appendix B. (Continued)

Data Summary Sheet #3: **Calculation of Imprecision: All Runs**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0

Source of pools: Controls
Analyte: BUN
Date/time:
Operator: RBP

	Low	Mid	High
(*R*) Pooled within-run variance†	0.133	3.53	2.60
(*S*) Variance of daily means‡	0.056	1.50	0.63
(*T*) Adjusted between-day variance§ (*S*) - (*R*)/3	0.011	0.323	0.0
(*U*) Total variance (*R*) + (*T*)	0.144	3.85	2.60
(*V*) Total standard deviation = $\sqrt{U}$	0.380	1.96	1.61
Grand mean value (from Worksheet #2)	9.0	54.7	92.6
(*W*) Total *CV*% = (*V*)/*Grand Mean Value* • 100%	4.22	3.59	1.74
Your allowable imprecision *CV*%	8%	3%	2%
Accept or reject	*Accept*	*Reject*	*Accept*

* BUN, blood urea nitrogen
†Because all days have the same number of data points, it is permissible to simply take the mean within-run variance for all accepted runs at each level.
‡Calculated as the variance of daily means from Data Summary Sheet #2 for each level.
§If less than zero, set equal to zero.

Appendix B. (Continued)

Data Summary Sheet #4a: **Multiple Regression Calculations (One Data Sheet Per Run)**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 8 August 1988
Operator: RBP
Day#: 1

Scale factor: [*Label Value* (*Mid*) - *Label Value* (*Low*)] = 41.5

	Slope: Worksheet #1a				Carry-over: Worksheet #1a		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	92	• 139 =	12,788.00		92	• 26 =	2,392.00
2	9	• -96 =	-864.00		9	• 130 =	1,170.00
3	54	• 11 =	594.00		54	• -102 =	-5,508.00
4	56	• 8 =	448.00		56	• 8 =	448.00
5	10	•-117 =	-1,170.00		10	• -4 =	-40.00
6	9	•-126 =	-1,134.00		9	• -126 =	-1,134.00
7	92	• 100 =	9,200.00		92	• -126 =	-11,592.00
8	95	• 100 =	9,500.00		95	• 100 =	9,500.00
9	59	• -19 =	-1,121.00		59	• 94 =	5,546.00
		Total =	28,241.00			Total =	782.00
		Total/678 =	41.653			Total/678 =	1.153
			(slope, *B1*)				(carry-over, *B2*)

	Nonlinearity: Worksheet #1a				Drift: Worksheet #1a		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	92	• 87 =	8924.00		92	• -52 =	-4,784.00
2	9	• 96 =	864.00		9	• -34 =	-306.00
3	54	•-237 =	-12,798.00		54	• -22 =	-1,188.00
4	56	•-234 =	-13,104.00		56	• -16 =	-896.00
5	10	• 117 =	1170.00		10	• 8 =	80.00
6	9	• 126 =	1134.00		9	• 26 =	234.00
7	92	• 126 =	11,592.00		92	• 26 =	2,392.00
8	95	• 126 =	11,970.00		95	• 26 =	2,470.00
9	59	•-207 =	-4,071.00		59	• 38 =	2,242.00
		Total =	-12,213.00			Total =	244.00
		Total/678 =	-4.987			Total/678 =	0.360
			(nonlinearity, *B3*)				(linear drift/test, *B4*)

The above coefficients are multipliers derived from the experimental design and should not be changed.

* BUN, blood urea nitrogen

Appendix B. (Continued)

Data Summary Sheet #4b: **Multiple Regression Calculations (One Data Sheet Per Run)**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 9 August 1988
Operator: RBP
Day#: 2

Scale factor: [*Label Value* (*Mid*) - *Label Value* (*Low*)] = 41.5

	Slope: Worksheet #1b				Carry-over: Worksheet #1b		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	92	• 139 =	12,788.00		92	• 26 =	2,392.00
2	9	• -96 =	-864.00		9	• 130 =	1,170.00
3	54	• 11 =	594.00		54	• -102 =	-5,508.00
4	54	• 8 =	432.00		54	• 8 =	432.00
5	9	•-117 =	-1,053.00		9	• -4 =	-36.00
6	8	•-126 =	-1,008.00		8	• -126 =	-1,008.00
7	91	• 100 =	9,100.00		91	• -126 =	-11,466.00
8	92	• 100 =	9,200.00		92	• 100 =	9,200.00
9	56	• -19 =	-1,064.00		56	• 94 =	5,264.00
		Total =	28,125.00			Total =	440.00
		Total/678 =	41.482			Total/678 =	0.649
			(slope, *B1*)				(carry-over, *B2*)

	Nonlinearity: Worksheet #1b				Drift: Worksheet #1b		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	92	• 87 =	8004.00		92	• -52 =	-4,784.00
2	9	• 96 =	864.00		9	• -34 =	-306.00
3	54	•-237 =	-12,798.00		54	• -22 =	-1,188.00
4	54	•-234 =	-12,636.00		54	• -16 =	-864.00
5	9	• 117 =	1,053.00		9	• 8 =	72.00
6	8	• 126 =	1,008.00		8	• 26 =	208.00
7	91	• 126 =	11,466.00		91	• 26 =	2,366.00
8	92	• 126 =	11,562.00		92	• 26 =	2,392.00
9	56	•-207 =	-11,592.00		56	• 38 =	2,128.00
		Total =	3,039.00			Total =	24.00
		Total/678 =	-4.482			Total/678 =	0.035
			(nonlinearity, *B3*)				(linear drift/test, *B4*)

The above coefficients are multipliers derived from the experimental design and should not be changed.

* BUN, blood urea nitrogen

Appendix B. (Continued)

Data Summary Sheet #4c: **Multiple Regression Calculations (One Data Sheet Per Run)**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 10 August 1988
Operator: RBP
Day#: 3

Scale factor: [*Label Value* (*Mid*) - *Label Value* (*Low*)] = 41.5

	Slope: Worksheet #1c				Carry-over: Worksheet #1c		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	93	• 139 =	12,927.00		93	• 26 =	2,418.00
2	9	• -96 =	-864.00		9	• 130 =	1,170.00
3	54	• 11 =	594.00		54	• -102 =	-5,508.00
4	54	• 8 =	432.00		54	• 8 =	432.00
5	9	•-117 =	-1,053.00		9	• -4 =	-36.00
6	9	•-126 =	-1,134.00		9	• -126 =	-1,134.00
7	92	• 100 =	9,200.00		92	• -126 =	-11,592.00
8	96	• 100 =	9,600.00		96	• 100 =	9,600.00
9	58	• -19 =	-1,102.00		58	• 94 =	5,452.00
		Total =	28,600.00			Total =	802.00
		Total/678 =	42.183			Total/678 =	1.183
			(slope, *B1*)				(carry-over, *B2*)

	Nonlinearity: Worksheet #1c				Drift: Worksheet #1c		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	93	• 87 =	8,091.00		93	• -52 =	-4,836.00
2	9	• 96 =	864.00		9	• -34 =	-306.00
3	54	•-237 =	-12,798.00		54	• -22 =	-1,188.00
4	54	•-234 =	-12,636.00		54	• -16 =	-864.00
5	9	• 117 =	1,053.00		9	• 8 =	72.00
6	9	• 126 =	1,134.00		9	• 26 =	234.00
7	92	• 126 =	11,592.00		92	• 26 =	2,392.00
8	96	• 126 =	12,096.00		96	• 26 =	2,496.00
9	58	•-207 =	-12,006.00		58	• 38 =	2,204.00
		Total =	-2,610.00			Total =	204.00
		Total/678 =	-3.850			Total/678 =	0.0301
			(nonlinearity, *B3*)				(linear drift/test, *B4*)

The above coefficients are multipliers derived from the experimental design and should not be changed.

* BUN, blood urea nitrogen

Appendix B. (Continued)

Data Summary Sheet #4d: **Multiple Regression Calculations (One Data Sheet Per Run)**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 11 August 1988
Operator: RBP
Day#: 4

Scale factor: [*Label Value* (*Mid*) - *Label Value* (*Low*)] = 41.5

	Slope: Worksheet #1d				Carry-over: Worksheet #1d		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	90	• 139 =	12,510.00		90	• 26 =	2,340.00
2	9	• -96 =	-864.00		9	• 130 =	1,170.00
3	54	• 11 =	594.00		54	• -102 =	-5,508.00
4	55	• 8 =	440.00		55	• 8 =	44.00
5	9	•-117 =	-1,053.00		9	• -4 =	-36.00
6	9	•-126 =	-1,134.00		9	• -126 =	-1,134.00
7	92	• 100 =	9,200.00		92	• -126 =	-11,592.00
8	94	• 100 =	9,400.00		94	• 100 =	9,400.00
9	52	• -19 =	-988.00		52	• 94 =	4,888.00
		Total =	28,105.00			Total =	-32.00
		Total/678 =	41.453			Total/678 =	-0.047
			(slope, *B1*)				(carry-over, *B2*)

	Nonlinearity: Worksheet #1d				Drift: Worksheet #1d		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	90	• 87 =	7,830.00		90	• -52 =	-4,680.00
2	9	• 96 =	864.00		9	• -34 =	-306.00
3	54	•-237 =	-12,798.00		54	• -22 =	-1,188.00
4	55	•-234 =	-12,870.00		55	• -16 =	-880.00
5	9	• 117 =	1,053.00		9	• 8 =	72.00
6	9	• 126 =	1,134.00		9	• 26 =	234.00
7	92	• 126 =	11,592.00		92	• 26 =	2,392.00
8	94	• 126 =	11,844.00		94	• 26 =	2,444.00
9	52	•-207 =	-10,764.00		52	• 38 =	1,976.00
		Total =	-2,115.00			Total =	64.00
		Total/678 =	-3.119			Total/678 =	0.094
			(nonlinearity, *B3*)				(linear drift/test, *B4*)

The above coefficients are multipliers derived from the experimental design and should not be changed.

* BUN, blood urea nitrogen

Appendix B. (Continued)

Data Summary Sheet #4e: **Multiple Regression Calculations (One Data Sheet Per Run)**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 15 August 1988
Operator: RBP
Day#: 6

Scale factor: [*Label Value* (*Mid*) - *Label Value* (*Low*)] = 41.5

	Slope: Worksheet #1e				Carry-over: Worksheet #1e		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	92	• 139 =	12,788.00		92	• 26 =	2,392.00
2	9	• -96 =	-864.00		9	• 130 =	1,170.00
3	52	• 11 =	572.00		52	• -102 =	-5,304.00
4	55	• 8 =	440.00		55	• 8 =	440.00
5	9	•-117 =	-1,053.00		9	• -4 =	-36.00
6	9	•-126 =	-1,134.00		9	• -126 =	-1,134.00
7	92	• 100 =	9,200.00		92	• -126 =	-11,592.00
8	94	• 100 =	9,400.00		94	• 100 =	9,400.00
9	53	• -19 =	-1,007.00		53	• 94 =	-4,982.00
		Total =	28,342.00			Total =	318.00
		Total/678 =	41.802			Total/678 =	0.469
			(slope, *B1*)				(carry-over, *B2*)

	Nonlinearity: Worksheet #1e				Drift: Worksheet #1e		
	Data	**Coefficient**	**Subtotal**		**Data**	**Coefficient**	**Subtotal**
1	92	• 87 =	8,004.00		92	• -52 =	-4,784.00
2	9	• 96 =	864.00		9	• -34 =	-306.00
3	52	•-237 =	-12,324.00		52	• -22 =	-1,144.00
4	55	•-234 =	-12,870.00		55	• -16 =	-880.00
5	9	• 117 =	1,053.00		9	• 8 =	72.00
6	9	• 126 =	1,134.00		9	• 26 =	234.00
7	92	• 126 =	11,592.00		92	• 26 =	2,392.00
8	94	• 126 =	11,844.00		94	• 26 =	2,444.00
9	53	•-207 =	-10,971.00		53	• 38 =	2,014.00
		Total =	-1,674.00			Total =	42.00
		Total/678 =	-2.469			Total/678 =	0.062
			(nonlinearity, *B3*)				(linear drift/test, *B4*)

The above coefficients are multipliers derived from the experimental design and should not be changed.

* BUN, blood urea nitrogen

Appendix B. (Continued)

Data Summary Sheet #5a: *t*-Statistic for Regression Coefficients: Single Run

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 8 August 1988
Operator: RBP
Day#: 1

(1) Compute intercept: B_0 = 52.89 (Average of all data.)
(2) Compute standard error or estimate: Use multiple linear regression equation below to calculate Y_j^* (predicted value of Y)

Assay #	Assay Transformed Value	Observed Value	Computed Y	Residual	Residual Squared
j	x_j	y_j	Y_j^*	$y_j - Y_j^*$	$(y_j - Y_j^*)^2$
0	0	51			
1	1	92	91.44	0.56	0.31
2	-1	9	9.65	-0.65	0.42
3	0	54	54.34	-0.34	0.12
4	0	56	55.85	0.15	0.02
5	-1	10	9.57	0.43	0.18
6	-1	9	8.78	0.22	0.05
7	1	92	92.45	-0.45	0.20
8	1	95	95.11	-0.11	0.01
9	0	59	58.81	0.19	0.04
				$S(y_j - Y_j^*)^2$=	1.35
				$S_{y \cdot x}$ =	0.58

$$S_{y \cdot x} = \sqrt{\frac{\Sigma\left(y_j - Y_j^*\right)^2}{4}}$$

$Y_j^* = B_0 + B_1 \bullet x_j + B_2 \bullet x_{j-1} + B_3(x_j^2 - 2/3) + B_4 \bullet t$ (see Appendix C) where x_j is the transformed value for x (i.e., low = -1, mid = 0, and high = +1) and the t multiplied by B_4 is time (-4 through +4).

(3) *t*-Statistics

(A) Regression Parameters	(B) Adjusted Regression Value†	(C) Standard Error Value	(D) Standard Error	(E) *t*-Statistic $t = B/D$
B_0	2.20	$S_{y \cdot x} \bullet 0.3333$ =	0.194	11.3
B_1	1.004	$S_{y \cdot x} \bullet 0.4135$ =	0.0058	0.69
B_2	2.77	$S_{y \cdot x} \bullet 0.4135$ =	0.240	4.80
B_3	-0.00290	$S_{y \cdot x} \bullet 0.7099$ =	0.000239	-12.09
B_4	0.360	$S_{y \cdot x} \bullet 0.1330$ =	0.077	4.65

t for 4 degrees of freedom is significant ($p < 0.01$) if $t > 4.6$ or $t < -4.6$.

* BUN, blood urea nitrogen
† Adjusted Regression Values: B_1adj = B_1/scale factor. Scale Factor = mid ref. – low ref. Use (B_1adj-1) for *t*-test. B_0adj = B_0 - (B_1 adj•mid ref.) = B_0(B_1 adj •50.5). Use B_0 adj for *t*-test. B_2adj = (B_2/ B_1)•100. Use B_2 for *t*-test. B_3adj = B_3/(scale factor • scale factor). Use B_3adj for *t*-test. B_4adj = B_4. (The standard error for B_1 is adjusted multiplying the result of C by 1/41.5). (The standard error for B_3 is adjusted multiplying the result of C by 1/[41.5 • 41.5]).

Appendix B. (Continued)

Data Summary Sheet #5b: *t*-Statistic for Regression Coefficients: Single Run

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 9 August 1988
Operator: RBP
Day#: 2

(1) Compute intercept: B_0 = 51.67 (Average of all data.)
(2) Compute standard error or estimate: Use multiple linear regression equation below to calculate Y_j^* (predicted value of *Y*)

Assay #	Assay Transformed Value	Observed Value	Computed *Y*	Residual	Residual Squared
j	x_j	y_j	Y_j^*	$y_j - Y_j^*$	$(y_j - Y_j^*)^2$
0	0	51			
1	1	92	91.51	0.49	0.24
2	-1	9	9.23	-0.23	0.05
3	0	54	53.94	0.06	0.00
4	0	54	54.62	-0.62	0.38
5	-1	9	8.69	0.31	0.10
9	-1	8	8.08	-0.08	0.01
7	1	91	91.08	-0.08	0.01
8	1	92	92.41	-0.41	0.017
9	0	56	55.45	0.55	0.03
				$S(y_j - Y_j^*)^2$=	1.26
				$S_{y \cdot x}$ =	0.56

$$S_{y \cdot x} = \sqrt{\frac{\Sigma\left(y_j - Y_j^*\right)^2}{4}}$$

$Y_j^* = B_0 + B_1 \bullet x_j + B_2 \bullet x_{j-1} + B_3(x_j^2 - 2/3) + B_4 \bullet t$ (see Appendix C) where x_j is the transformed value for *x* (i.e., low = -1, mid = 0, and high = +1) and the *t* multiplied by B_4 is time (-4 through +4).

(3) *t*-Statistics

(A) Regression Parameters	(B) Adjusted Regression Value†	(C) Standard Error Value	(D) Standard Error	(E) *t*-Statistic *t = B/D*
B_0	1.19	$S_{y \cdot x} \bullet 0.3333$ =	0.187	6.36
B_1	1.000	$S_{y \cdot x} \bullet 0.4135$ =	0.0055	0.00
B_2	1.56	$S_{y \cdot x} \bullet 0.4135$ =	0.23	2.82
B_3	-0.00260	$S_{y \cdot x} \bullet 0.7099$ =	0.000232	-11.20
B_4	0.035	$S_{y \cdot x} \bullet 0.1330$ =	0.07	0.50

t for 4 degrees of freedom is significant ($p < 0.01$) if $t > 4.6$ or $t < -4.6$.

* BUN, blood urea nitrogen
† Adjusted Regression Values: B_1adj = B_1/scale factor. Scale Factor = mid ref. – low ref. Use (B_1adj-1) for *t*-test. B_0adj = B_0 - (B_1 adj•mid ref.) = B_0(B_1 adj •50.5). Use B_0 adj for *t*-test. B_2adj = (B_2/ B_1)•100. Use B_2 for *t*-test. B_3adj = B_3/(scale factor • scale factor). Use B_3adj for *t*-test. B_4adj = B_4. (The standard error for B_1 is adjusted multiplying the result of C by 1/41.5). (The standard error for B_3 is adjusted multiplying the result of C by 1/[41.5 • 41.5]).

Appendix B. (Continued)

Data Summary Sheet #5c: *t*-Statistic for Regression Coefficients: Single Run

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 10 August 1988
Operator: RBP
Day#: 3

(1) Compute intercept: B_0 = 52.67 (Average of all data.)
(2) Compute standard error or estimate: Use multiple linear regression equation below to calculate Y_j^* (predicted value of Y)

Assay #	Assay Transformed Value	Observed Value	Computed Y	Residual	Residual Squared
j	x_j	y_j	Y_j^*	$y_j - Y_j^*$	$(y_j - Y_j^*)^2$
0	0	51			
1	1	93	92.36	0.64	0.41
2	-1	9	9.48	-0.48	0.23
3	0	54	53.45	0.55	0.30
4	0	54	54.93	-0.93	0.86
5	-1	9	9.20	-0.20	0.04
6	-1	9	8.32	0.68	0.46
7	1	92	92.99	-0.99	0.98
8	1	96	95.65	0.35	0.12
9	0	58	57.62	0.38	0.14
				$S(y_j - Y_j^*)^2$=	3.55
				$S_{y \cdot x}$ =	0.94

$$S_{y \cdot x} = \sqrt{\frac{\Sigma\left(y_j - Y_j^*\right)^2}{4}}$$

$Y_j^* = B_0 + B_1 \bullet x_j + B_2 \bullet x_{j-1} + B_3(x_j^2 - 2/3) + B_4 \bullet t$ (see Appendix C) where x_j is the transformed value for x (i.e., low = -1, mid = 0, and high = +1) and the t multiplied by B_4 is time (-4 through +4).

(3) *t*-Statistics

(A) Regression Parameters	(B) Adjusted Regression Value†	(C) Standard Error Value	(D) Standard Error	(E) *t*-Statistic *t = B/D*
B_0	1.34	$S_{y \cdot x} \bullet 0.3333$ =	0.314	4.27
B_1	1.016	$S_{y \cdot x} \bullet 0.4135$ =	0.0094	1.70
B_2	2.80	$S_{y \cdot x} \bullet 0.4135$ =	0.39	3.03
B_3	-0.00224	$S_{y \cdot x} \bullet 0.7099$ =	0.000389	-5.75
B_4	0.301	$S_{y \cdot x} \bullet 0.1330$ =	0.13	2.32

t for 4 degrees of freedom is significant ($p < 0.01$) if $t > 4.6$ or $t < -4.6$.

* BUN, blood urea nitrogen
† Adjusted Regression Values: B_1adj = B_1/scale factor. Scale Factor = mid ref. – low ref. Use (B_1adj-1) for *t*-test. B_0adj = B_0 - (B_1 adj•mid ref.) = B_0(B_1 adj •50.5). Use B_0 adj for *t*-test. B_2adj = (B_2/ B_1)•100. Use B_2 for *t*-test. B_3adj = B_3/(scale factor • scale factor). Use B_3adj for *t*-test. B_4adj = B_4. (The standard error for B_1 is adjusted multiplying the result of C by 1/41.5). (The standard error for B_3 is adjusted multiplying the result of C by 1/[41.5 • 41.5]).

Appendix B. (Continued)

Data Summary Sheet #5d: *t*-Statistic for Regression Coefficients: Single Run

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1
Source of pools: Controls
Analyte: BUN
Date/time: 11 August 1988
Operator: RBP
Day#: 4

(1) Compute intercept: B_0 = 51.56 (Average of all data.)
(2) Compute standard error or estimate: Use multiple linear regression equation below to calculate Y_j^* (predicted value of *Y*)

Assay #	Assay Transformed Value	Observed Value	Computed *Y*	Residual	Residual Squared
j	x_j	y_j	Y_j^*	$y_j - Y_j^*$	$(y_j - Y_j^*)^2$
0	0	51			
1	1	90	91.59	-1.59	2.53
2	-1	9	8.73	0.27	0.07
3	0	54	53.49	0.51	0.26
4	0	55	53.54	1.46	2.13
5	-1	9	9.06	-0.06	0.00
6	-1	9	9.20	-0.20	0.04
7	1	92	92.20	-0.20	0.04
8	1	94	92.20	1.80	3.24
9	0	52	53.97	-1.97	3.88
				$S(y_j - Y_j^*)^2$=	12.16
				$S_{y \cdot x}$ =	1.74

$$S_{y \cdot x} = \sqrt{\frac{\Sigma\left(y_j - Y_j^*\right)^2}{4}}$$

$Y_j^* = B_0 + B_1 \bullet x_j + B_2 \bullet x_{j-1} + B_3(x_j^2 - 2/3) + B_4 \bullet t$ (see Appendix C) where x_j is the transformed value for *x* (i.e., low = -1, mid = 0, and high = +1) and the *t* multiplied by B_4 is time (-4 through +4).

(3) *t*-Statistics

(A) Regression Parameters	(B) Adjusted Regression Value†	(C) Standard Error Value	(D) Standard Error	(E) *t*-Statistic *t = B/D*
B_0	1.11	$S_{y \cdot x} \bullet 0.3333$ =	0.58	1.91
B_1	0.999	$S_{y \cdot x} \bullet 0.4135$ =	0.017	-0.059
B_2	-0.11	$S_{y \cdot x} \bullet 0.4135$ =	0.72	-0.07
B_3	-0.00181	$S_{y \cdot x} \bullet 0.7099$ =	0.000720	-2.52
B_4	0.094	$S_{y \cdot x} \bullet 0.1330$ =	0.23	0.41

t for 4 degrees of freedom is significant ($p < 0.01$) if $t > 4.6$ or $t < -4.6$.

* BUN, blood urea nitrogen
† Adjusted Regression Values: B_1adj = B_1/scale factor. Scale Factor = mid ref. – low ref. Use (B_1adj-1) for *t*-test. B_0adj = B_0 - (B_1 adj•mid ref.) = B_0(B_1 adj •50.5). Use B_0 adj for *t*-test. B_2adj = (B_2/ B_1)•100. Use B_2 for *t*-test. B_3adj = B_3/(scale factor • scale factor). Use B_3adj for *t*-test. B_4adj = B_4. (The standard error for B_1 is adjusted multiplying the result of C by 1/41.5). (The standard error for B_3 is adjusted multiplying the result of C by 1/[41.5 • 41.5]).

Appendix B. (Continued)

Data Summary Sheet #5e: *t*-Statistic for Regression Coefficients: Single Run

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0
Run#: 1

Source of pools: Controls
Analyte: BUN
Date/time: 15 August 1988
Operator: RBP
Day#: 6

(1) Compute intercept: B_0 = 51.67 (Average of all data.)
(2) Compute standard error or estimate: Use multiple linear regression equation below to calculate Y_j^* (predicted value of Y)

Assay #	Assay Transformed Value	Observed Value	Computed Y	Residual	Residual Squared
j	x_j	y_j	Y_j^*	$y_j - Y_j^*$	$(y_j - Y_j^*)^2$
0	0	52			
1	1	92	92.40	-0.40	0.16
2	-1	9	9.34	-0.32	0.10
3	0	52	52.72	-0.72	0.52
4	0	55	53.25	1.75	3.06
5	-1	9	9.04	-0.04	0.00
6	-1	9	8.63	0.37	0.14
7	1	92	92.30	-0.30	0.09
8	1	94	93.30	0.70	0.49
9	0	53	54.03	-1.03	1.06
				$S(y_j - Y_j^*)^2$=	5.62
				$S_{y \cdot x}$ =	1.18

$$S_{y \cdot x} = \sqrt{\frac{\Sigma\left(y_j - Y_j^*\right)^2}{4}}$$

$Y_j^* = B_0 + B_1 \bullet x_j + B_2 \bullet x_{j-1} + B_3(x_j^2 - 2/3) + B_4 \bullet t$ (see Appendix C) where x_j is the transformed value for x (i.e., low = -1, mid = 0, and high = +1) and the t multiplied by B_4 is time (-4 through +4).

(3) *t*-Statistics

(A) Regression Parameters	(B) Adjusted Regression Value†	(C) Standard Error Value	(D) Standard Error	(E) *t*-Statistic $t = B/D$
B_0	0.80	$S_{y \cdot x} \bullet 0.3333$ =	0.393	2.04
B_1	1.007	$S_{y \cdot x} \bullet 0.4135$ =	0.012	0.58
B_2	1.12	$S_{y \cdot x} \bullet 0.4135$ =	0.49	0.96
B_3	-0.00143	$S_{y \cdot x} \bullet 0.7099$ =	0.000488	-2.94
B_4	0.062	$S_{y \cdot x} \bullet 0.1330$ =	0.16	0.39

t for 4 degrees of freedom is significant ($p < 0.01$) if $t > 4.6$ or $t < -4.6$.

* BUN, blood urea nitrogen
† Adjusted Regression Values: B_1adj = B_1/scale factor. Scale Factor = mid ref. – low ref. Use (B_1adj-1) for *t*-test. B_0adj = B_0 - (B_1 adj•mid ref.) = B_0(B_1 adj •50.5). Use B_0 adj for *t*-test. B_2adj = (B_2/ B_1)•100. Use B_2 for *t*-test. B_3adj = B_3/(scale factor • scale factor). Use B_3adj for *t*-test. B_4adj = B_4. (The standard error for B_1 is adjusted multiplying the result of C by 1/41.5). (The standard error for B_3 is adjusted multiplying the result of C by 1/[41.5 • 41.5]).

Appendix B. (Continued)

Data Summary Sheet #6: **Multiple Regression Summary: All Runs**

Method/instrument: BUN* Analyzer
Assigned Values (C):
Low: 9.0
Mid: 50.5
High: 92.0

Source of pools: Controls
Analyte: BUN
Date/time: 15 August 1998
Operator: RBP

Day	Intercept (B_0adj)	Slope (B_1adj)	%Carry-over (B_2adj)	Nonlinearity (B_3adj)	Drift (B_4)	$S_{y \cdot x}$
1 Value	2.20	1.004	2.77	-0.00290	0.36	0.58
t	11.3	0.69	4.80	-12.09	4.65	
2 Value	1.19	1.000	1.56	-0.00260	0.035	0.56
t	6.36	0.000	2.82	-11.20	0.50	
3 Value	1.34	1.016	2.80	-0.00224	0.301	0.94
t	4.27	1.70	3.03	-5.75	2.32	
4 Value	1.11	0.999	-0.11	-0.00181	0.094	1.74
t	1.91	-0.059	-0.07	-2.52	0.41	
5 Value	0.80	1.007	1.12	-0.00143	0.062	1.18
t	2.04	0.58	0.96	-2.94	0.39	

t-value comes from column E in Data Summary Sheet #5.

Summary of Runs
using a one sample sign test:

The following table shows the average of each parameter estimate and whether each parameter estimate is significant according to a one sample sign test. If all values of a parameter have the same sign (1 is subtracted for the slope values) then the sign test is significant at the p = 0.06 level.

	Intercept (B_0adj)	Slope (B_1adj)	%Carry-over (B_2adj)	Nonlinearity (B_3adj)	Drift (B_4)	$S_{y \cdot x}$
Value	1.33	1.005	1.63	-0.0022	0.171	1.09
significant?	Yes	Yes	No	Yes	Yes	

Conclusions: *Nonlinearity is different from zero (this may also be observed by looking at the difference plot). With significant nonlinearity, the slope† and intercept parameters may not be meaningful when considered as measures of proportional and constant error. The percent sample carry-over is not detectably different from zero. The drift is detectably different from zero but very small. There was one unexplained outlier that had been deleted from the analysis. $S_{y \cdot x}$ a measure of imprecision across all levels, is similar to the individual imprecision estimates by level in Data Sheet #3.*

* BUN, blood urea nitrogen
† The slope was tested as slope-1, which represents a test for slope bias with an expected value of zero.

Appendix B. (Continued)

Data Summary Sheet #7: **Individual Patient Sample Comparisons (Optional)**

Device X (old): BUN* Analyzer Analyte: BUN*
Device Y (new): BUN* Analyzer

	Low Concentration			Medium Concentration			High Concentration		
Day	***Y***	***X***	**(Bias) *Y-X***	***Y***	***X***	**(Bias) *Y-X***	***Y***	***X***	**(Bias) *Y-X***
1	10	9	1	45	42	3	98	98	0
2	11	11	0	52	48	4	99	98	1
3	9	10	-1	56	51	5	98	98	0
4	9	9	0	55	51	4	87	85	2
5	10	10	0	46	43	3	82	82	0
Average Concentration of X =									
		9.8			47			92.2	
Average Bias =									
			0.0			3.8			0.6
% Bias = (Average Bias/Average Concentration) 100 =									
		0			8.1%			0.7%	

Comments:
The analyst was easily able to choose samples very near the mean of the low pool and used a different sample each day. Samples near the mean of the medium concentration were more rare, but still a different sample was used each day. For the high samples, not many at all were available, and it was necessary to use the same sample for the first three days; then two others were submitted for analysis on days 4 and 5. The analyst knew that BUN is relatively stable in serum and kept the high sample refrigerated between uses. The same nonlinearity observed in the control pools was observed in the patient samples, suggesting that it was not due to a matrix effect. The presence of an unacceptable bias in the middle range requires correction before the method is accepted for routine use.

* BUN, blood urea nitrogen

Appendix C. Statistical Explanation

C1. Computing Components of Variance

The formula for the uncorrected total standard deviation is as follows:

$$SD_L = \sqrt{\frac{\sum_{i=1}^{N_L}\left(x_i - \bar{x}\right)^2}{N_L - 1}} \quad (1)$$

where

L = level number (1 = low, 2 = mid, 3 = high)
N_L = total number of observations at level L
x_i = the I^{th} point at level L ($I = 1,...,N_L$); and,
$\bar{x}$ = the mean of the points at level L.

The three components should be calculated separately for each level, as follows:

Equation Description	$J = 1$	$J > 1$	Equation #
within-run ("uncorrected")	$S_w^2 = \frac{\sum_i \sum_k \left(x_{ik} - \bar{x}_i\right)^2}{I(K-1)}$	$S_w^2 = \frac{\sum_i \sum_j \sum_k \left(x_{ijk} - \bar{x}_{ij}\right)^2}{IJ(K-1)}$	(2)
between-run ("uncorrected")	Not Applicable	$S_R^2 = \frac{\sum_i \sum_j \left(\bar{x}_{ij} - \bar{x}_i\right)^2}{I(J-1)}$	(3)
between-day ("uncorrected")	$S_D^2 = S_R^2 = \frac{\sum_i \left(\bar{x}_{i-} \bar{x}\right)^2}{I-1}$	$S_D^2 = \frac{\sum_i \left(\bar{x}_i - \bar{x}\right)^2}{I-1}$	(4)
within-run estimate	$\hat{o}_w^2 = S_w^2$	$\hat{o}_w^2 = S_w^2$	(5)
between-run estimate	(Not Used)	$\hat{o}_R^2 = S_R^2 - \frac{S_w^2}{K}$	(6)
between-day estimate	$\hat{\sigma}_D^2 = \hat{\sigma}_R^2 = S_D^2 - \frac{S_W^2}{K} = S_R^2 - \frac{S_W^2}{K}$ (actual)	$\hat{\sigma}_D^2 = S_D^2 \frac{\hat{\sigma}_R^2}{J} - \frac{S_W^2}{JK} = S_D^2 - \frac{S_R^2}{J}$	(7)
total estimate	$\hat{o}_T^2 = \hat{o}_D^2 + \hat{o}_w^2$	$\hat{o}_T^2 = \hat{o}_D^2 + \hat{o}_R^2 + \hat{o}_w^2$	(8)

Note: If either equation 6 or 7 results in a negative number, set that component equal to zero.

Where:

I = number of days (usually five);
J = number of runs on each day (one in this document);
K = number of observations in each run (i.e., three at each level);
x_{ijk} = observation k in run j on day I;
$\bar{x}_{Ij}$ = mean of the K observations in run j on day I
$\bar{x}_I$ = mean of all observations on day I;
$\bar{x}$ = grand mean of all observations.

Remember that these formulae are used at each concentration level separately.

Appendix C. (Continued)

C2. Multiple Regression Procedure for Sources of Unacceptable Imprecision

For calculation purposes, each "level" of the variables included in the model is represented by a dummy variable value. The dummy variables used are:

Concentration Level	(-1, 0, +1)	(for slope)
Prior Sample Concentration (*conc.*)	(-1, 0, +1)	(for carry-over)
Concentration Squared	(0, +1)	(for linearity)
Time	(coded -4, -3,...,2, 3, 4)	(for linear drift)

As an example, the table below represents the coding (or conversion of the label values) for the data from the run on the first sheet in Appendix B:

	Dependent Variable	Independent Variables			
	Concentration	Prior Concentration	Concentration^2	Time	
	x_j	x_{j-1}	$(x_j)^2 - 2/3$	t_j	Result
1	+1	0	0.333	-4	92
2	-1	+1	0.333	-3	9
3	0	-1	-0.667	-2	54
4	0	0	-0.667	-1	56
5	-1	0	0.333	0	10
6	-1	-1	0.333	+1	9
7	+1	-1	0.333	+2	92
8	+1	+1	0.333	+3	95
9	0	+1	-0.667	+4	59

The "0" in the "Prior Concentration" column for the first sample comes from the midlevel priming sample, which precedes the sample sequence. Time (*t*), used for the assessment of within-run (linear) drift, is coded from -4 to +4 instead of 1 to 9 for calculation purposes.

$$Y_j = B_o + B_1 \cdot x_j + B_2 \cdot x_{j-1} + B_3 (x_j^2 \cdot \frac{2}{3}) + B_4 \cdot t \qquad (9)$$

The model that represents the evaluation sequence may be written as follows:

where j = 1 to 9.

In this model equation, x_j represents the concentration (-1, 0, or +1) of sample j, x_{j-1} is the concentration of the immediately preceding sample, and *t* represents time (coded -4 to +4). Y_j is the predicted result for sample *j*.

Multiple linear regression should be used to estimate the coefficients in the model. However, since the independent variables take on the same values for every run, the calculations can be greatly simplified. Taking advantage of the efficiency of this design,[*] the vectors of multipliers may be predefined as seen in Data Summary Sheet #4.

[*] Krouwer JS, Stewart WN, and Schlain B. A multifactor experimental design for evaluating random-access analyzers. *Clin. Chem*. 1988;33:18941896.

Appendix C. (Continued)

These numbers are used to produce the regression estimates of intercept, slope, carry-over, nonlinearity and drift, represented by B_0, B_1, B_2, B_3, and B_4, respectively. The calculations are illustrated in the appendix. The carry-over term reported in the summary has the following transformation for each estimate: percent carry-over=$(B_2/B_1) \bullet 100$. The slope and intercept terms (B_0,B_1) are adjusted as shown in Data Sheet #5 to give expected values of 1 and 0.

After obtaining these regression coefficient estimates, the variances and *t*-statistic values should be derived as illustrated in Appendix B on Data Summary Sheet #4. The calculations can be done fairly easily with a hand calculator for Data Summary Sheets #3 and #4.

If any of the *t* statistics for drift, carry-over, or nonlinearity are found significant, it should be ascertained whether the same problem recurs in other runs.

Additionally, the parameter estimates are used as data for a t-test to determine if a parameter estimate is different from zero across all runs. If so, then the system/method should be evaluated more extensively to determine the actual source of the problem, or the manufacturer should be contacted. Usually however, if the problem appears to be limited only to one run, it may be safely ignored.

NCCLS consensus procedures include an appeals process that is described in detail in Section 9.0 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

EP10-A: *Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline*

Appendix B

1. We believe that the value for "Nonlinearity (B_3)" should be –2.19e-3 (versus the published –3.78) which is –3.78 divided by the square root of the scale factor (41.5).

- **In the summary of runs on page 35, the document had switched back to the unadjusted coefficient B3. The subcommittee agrees with the commenter and the nonlinearity term description has been changed from B3 to B3adj and the value has been changed to –0.0022 (the rounded value from the commenter). Note that the square of the scale factor is used, not the square root.**

2. We believe that the value for "Drift (B_4)" should be 0.171 (versus the published value of 1.71).

- **The subcommittee thanks the commenter for pointing out this error. The document has been corrected.**

3. We are unable to figure out how the t-values listed in this table were calculated. For example, for the calculation of Drift (B_4), the approach suggested is:

$$D = Sy.x \bullet 0.1330$$
$$D = 1.09 \bullet 0.1330 = 0.14497$$

$$t = \frac{B_4}{D}$$
$$t = \frac{0.171}{0.14497} = 1.18 \ (\textit{versus the published } 2.56).$$

- **The calculation of t-values has caused some confusion. To explain what was done previously, each set of parameter estimates for the five runs was considered to be an independent observation. For the drift example, using the parameter estimate from each run gave an average of 0.171 and a standard deviation of 0.149. The t-value = 0.171 / (0.149 / sqrt(5)) = 2.57. However, the document has been changed (see Section 13.2) and this *t*-test has been deleted. The recommended summary method is now to use a one-sample sign test.**

Appendix C

4. Appendix C explains a two-stage nested design with "Day" and "Run" as two factors. There are I levels of factor "Day." J levels of factor "Run" nested under each level of "Day." This kind of design

is also discussed in Section 13-2 of "Design and Analysis of Experiments" by Douglas C. Montgomery, 3rd edition, pages 440 to 450. In our opinion:

a. Formula 4 under J=1 should be:

$$S_D^2 = S_R^2 = \frac{\sum_i \left(\bar{x}_{i-} \bar{x}\right)^2}{I - 1}$$

b. Formula 4 under J>1 should be:

$$S_D^2 = \frac{\sum_i \left(\bar{x}_i - \bar{x}\right)^2}{I - 1}$$

c. Formula 7 under J=1 should be:

$$\hat{\sigma}_D^2 = \hat{\sigma}_R^2 = S_D^2 - \frac{S_W^2}{K} = S_R^2 - \frac{S_W^2}{K}$$

d. Formula 7 under J>1 should be:

$$\hat{\sigma}_D^2 = S_D^2 \frac{\hat{\sigma}_R^2}{J} - \frac{S_W^2}{JK} = S_D^2 - \frac{S_R^2}{J}$$

There was a comment regarding formula 7 of Appendix C in NCCLS guideline EP10-T2. We believe the comment is correct; however, the response is not. The corrected formula 7 in item d above matches the formula in NCCLS guideline EP5-A. The problems in items a, b, and c may simply be typographical.

- **The subcommittee thanks the commenter for identifying these errors. The document has been modified accordingly.**

Summary of Delegate Comments and Subcommittee Responses

EP10-A2: *Preliminary Evaluation of Quantitative Clinical Laboratory Methods; Approved Guideline—Second Edition*

General

1. EP10 sends the wrong message, i.e., that patient results may be reported based on a "preliminary evaluation." This is unacceptable.

   CLIA requires that laboratories verify and/or validate measurement procedures before placing them in use. This requires more than a "preliminary evaluation." Patient results shall not be reported by a laboratory until the method is validated for its intended use.

   There is no reason that the concept used in other regulated (and non-regulated) industries should not apply to clinical laboratories, in partnership with their suppliers i.e., the IVD manufacturers that must validate their methods prior to placing them on the market.

   In other regulated industries, if the method is already validated as is the case with commercial IVD systems, the method can be transferred to the end-user lab for the same intended use by completing a transfer protocol provided by the transferring laboratory and showing that the results meet pre-determined acceptance criteria.

   The experiment described in EP10 meets the requirements for a transfer protocol. The changes that would be necessary are easily made.

   The NCCLS evaluation protocols were developed before quality systems were adopted, and they no longer meet the needs of clinical laboratories. The potential for use by manufacturers is acknowledged in each of the documents, but in practice they are not suitable for use without substantial modification. The QSR regulates the development of IVD assays and imposes design controls, which include design verification and design validation – activities that require scientifically valid evaluation protocols. Now that quality systems are used in laboratories as well as manufacturers, and NCCLS is committed to promote quality systems, the evaluation protocols should be reviewed and revised to fit into the new quality system structure that encompasses the joint partnership between clinical laboratories and manufacturers.

- **EP10 stresses that it is a preliminary evaluation and not a total means to determine whether to report patient results.**

  The commenter suggests that (all) evaluation protocols should serve as "transfer protocols" for validated assays to facilitate meeting regulations. NCCLS evaluation protocols are not written as part of regulations and NCCLS protocols have no regulatory standing. The commenter should review EP15, *User Demonstration of Performance for Precision and Accuracy* which is perhaps closest to what is being referred to as a "transfer protocol".

  The commenter implies that an assay "validated" by a manufacturer and then transferred to a laboratory with "transfer protocol" is sufficient to guarantee quality. Certainly, the amount of validation performed by manufacturers is vital. However, laboratories have a responsibility to provide quality results that goes beyond meeting regulations. There are many instances of assay problems (often published in *Clinical Chemistry*) that were discovered in laboratories in commercially available (e.g., "validated") assays. Evaluation protocols are tools that laboratories can use should they wish to examine various assay parameters related to quality.

The adoption of quality systems by laboratories does not enter into the usefulness of EP10. It is a scientifically valid protocol that has been in use for over 25 years.

The commenter suggests that protocols such as EP10 are not suitable for use by manufacturers without suitable modification. This is simply not true. EP10 was developed by a manufacturer (see Reference 1).

Foreword

2. The statement, "Before using a new method or instrument for in vitro diagnostic use, the laboratory must make a preliminary decision about its acceptability" is an unacceptable at a time when concern over medical errors has placed a spotlight on laboratories. It says that the laboratory may start using a new method or instrument before it is sure it is acceptable.

- **The commenter should note that the Foreword goes on to say, "Rather, this experiment is a quick check to rule out major problems and a starting point for accumulating data and experience that will enable the user to make a final decision. The primary purpose of this document is to help detect performance problems that would warrant immediate correction, referral to the manufacturer, or expanded investigation before a new device is placed into service."**

3. The statement, "This initial performance check is neither a rigorous characterization of long-term performance nor an evaluation of the many factors that can affect results produced by the device" implies that laboratories may start using a device before evaluating the many factors that can affect the results. It does not convey that the device has already been validated by the manufacturer for the intended use in the laboratory, and the laboratory simply needs to verify it is meeting the manufacturer's claims.

- **See response to Comment 2. The commenter's statement is a bit at odds with the statement above (from the same commenter) about concern for medical errors. That the manufacturer has validated a method is understood. However, it is also the responsibility of the laboratory, as much as possible, to prevent medical errors. This may go beyond simply verifying that the assay meets the manufacturer's claims. Validating a method is not the same as validating a particular instrument, as installed in a particular laboratory.**

4. The statement, "Rather, this experiment is a quick check to rule out major problems and a starting point for accumulating data and experience that will enable the user to make a final decision. The primary purpose of this document is to help detect performance problems that would warrant immediate correction, referral to the manufacturer, or expanded investigation before a new device is placed into service" is not clear. It seems to suggest it is only necessary to fix major problems before a device may be used, implying "minor" problems can be tolerated. What does this mean? In a quality system, the laboratory must provide objective evidence that the method is suitable for its intended use. This should be the purpose of EP10.

- **The commenter is putting too much burden on EP10. It is one in a series of evaluation protocols, all of which are designed to assess various aspects of assay quality. EP10 is not a "stand-alone" document to validate all aspects of an assay.**

5. The "Note on Terminology" suggests the terms listed are merely synonyms. This is not true the concepts are different. NCCLS's credibility will be harmed by such errors. Trueness is a component of accuracy. Accuracy is more than trueness. Method accuracy, which is measured as bias, is now

called trueness. Measurand is not another word for analyte. The analyte is the substance being measured – period. The measurand is the analyte, along with the type of quantity (concentration, amount, etc.,), the sample matrix (serum, blood, water), and the units (IS, mg/dL, etc.). To make statements suggesting these are just synonyms trivializes the importance of what the standards writers are trying to accomplish, and undermines NCCLS's credibility with the rest of the world.

- **The terms have been reexamined and ISO definitions used. Also, the "Note on Terminology" has been expanded to respond to the commenter's concerns. Also, please see the response to Comment 13.**

6. The "introduction" to the quality system approach suggests that NCCLS invented quality systems. There is no reference to the ISO quality system standard.

- **The approach is based on the model presented in the most current edition of NCCLS HS1—*A Quality System Model for Health Care.* As stated in HS1, the approach applies concepts of quality design that are consistent with those described in the ISO 9000 series of standards for quality management.**

Introduction

7. We recommend modifying the first two sentences in the first paragraph of the Introduction to read, "This document describes a transfer protocol for use when implementing a new method that has been previously validated by the manufacturer. The protocol evaluates linearity, proportional and constant bias, linear drift, sample carry-over, and precision against acceptance criteria based on the manufacturer's claims. The protocol can also be used to verify performance after a minor modification has been introduced."

- **It is not exactly clear what the commenter means by a "transfer protocol." The commenter implies that the user has knowledge of each manufacturer's claim and the way it was evaluated. This may not be the case, especially for systematic effects such as drift and carryover. The manufacturer may not even mention these terms in product labeling.**

8. We recommend modifying the second paragraph of the Introduction to read, "The experiment is intended to provide initial estimates of performance characteristics that will be monitored by the laboratory's routine quality control system."

- **The laboratory quality control system is routinely carried out by assaying one (commercial control) sample per shift and will not permit evaluation of the same quantities that are estimated by the EP10 protocol. However, there is nothing to preclude the user from running EP10 after a method is in use, which will allow the user to monitor these performance estimates.**

Scope

9. "This document describes a protocol that may be used to transfer a validated method to an end-user laboratory. The acceptance criteria are based on the performance claims of the manufacturer. The protocol may also be used to verify that performance is suitable for use when a minor modification is introduced. The protocol is not intended for the full validation of a measurement procedure."

- **See response to Comment 7.**

Definitions

10. In the definition of "accuracy," delete "{/measurand}." It is incorrect.

- **The definition has been modified for consistency with ISO 3534-1*, Statistics—Vocabulary and Symbols—Part I: Probability and general statistical terms.***

11. Even though the term "measurand" has been adopted, the term "analyte" still appears in the text (e.g., in the definition of carryover, the beginning of Section 5 on Materials).

- **NCCLS is phasing in international terms. Hence there will be some overlap.**

12. Define the term "analyte." It is not the same as measurand.

- **The definition of analyte has been for consistency with ISO 18153, *In vitro diagnostic medical devices—Measurement of quantities in samples of biological origin—Metrological traceability for catalytic concentration of enzymes assigned to calibrators and control materials*.**

13. In the definition of "measurand," eliminate indication that "analyte" is a synonym. Also, we recommend using the ISO definition.

- **The use of parentheses in the definitions section is according to NCCLS style as outlined in document NRSCL8—*Terminology and Definitions for Use in NCCLS Documents*, and indicates that the parenthetical term is deprecated and its use is discouraged.**

14. Delete the definition of "preliminary evaluation."

- **This change has been made as recommended.**

15. In the definition of the term "reproducibility," the phrase "under changed conditions of measurement" defines Intermediate Precision, when some but not all variables are expressed. We recommend using the ISO 5275 definition.

- **The definition has been modified for consistency with ISO 3534-1*, Statistics—Vocabulary and Symbols—Part I: Probability and general statistical terms.***

16. In the definition of "trueness," eliminate indication that "accuracy" is a synonym.

- **See the response to Comment 13.**

Section 10.2

17. The first paragraph does not make the point that no data (even outliers) may be discarded. The purpose for identifying outliers is so they will not be used in calculating estimates of central tendency or average dispersion, which would be distorted by including values that are not part of the same statistical population. This does not mean they can be ignored or discarded. Outliers represent important information about the performance of the device. The essence of this comment should be included.

- **Outliers may be discarded (see paragraph 2 of Section 10.2). The commenter is correct in stressing the importance of outliers. However, this is stated in the second paragraph ("If an outlier is found, every effort should be made to determine the cause; it may indicate a fundamental problem.")**

18. Treatment of outliers in the statement, "A single run that has an outlier may be replaced with another run," is far too casual. If an outlier occurs in such a small data set, this may be indicative of a serious problem. It should be followed by more extensive testing, not just substitute another run.

- **Although the commenter is correct about the seriousness of outliers, one option stated in this section is "terminate this preliminary investigation and begin an expanded evaluation of imprecision."**

Section 11.1

19. In the first sentence, a transfer protocol requires predetermined acceptance criteria, not goals.

- **The subcommittee does not understand the difference between "goals" and "predetermined acceptance criteria;" therefore, the text has been maintained.**

20. As stated in the seventh sentence, goals may be beyond the scope; however, determining acceptance criteria must be part of this document.

- **The commenter is correct in principle. However, to provide goals for a wide range of assays for each performance parameter is truly beyond the scope of this document. Also, see response to Comment 19.**

21. Delete the reference to EP11 as it is not relevant.

- **The subcommittee believes the reference to EP11 may be helpful.**

22. Modify the last sentence to read, "To help define imprecision goals, one could use the manufacturer's labeling claims or performance of existing or similar analytes (measurands)."

- **The use of manufacturer's labeling claims has been added.**

Section 12.3

23. Goals are not used in validation testing; predetermined acceptance criteria are needed. The results must be compared to the acceptance criteria to determine pass/fail.

- **See response to Comment 19.**

Section 15

24. "The subcommittee changed . . ." does not belong in a guideline. Just state the requirement.

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

25. The final decision should be based on the claims (which are specifications). The specifications should have been compared to medical requirements sand found acceptable before this.

- **The subcommittee believes the current sentence is appropriate. "The final decision as to the acceptability of the method should be based on the medical usefulness of the assay results and currently accepted standards of laboratory practice." The commenter's remark is of course correct but obvious. If a manufacturer's claims were unacceptable with respect to medical requirements, then the assay should not be evaluated at all.**

References

26. The references are incomplete. Cite ISO standards and test method validation literature (e.g., DeSain).

- **Documentation for GMP and ISO 9000 is beyond the scope of this guideline; therefore, the recommended references have not been included.**

Appendix A

27. If possible, the worksheet for Appendix A should be made as an "Excel" file so that laboratorians may enter the values, and the calculations be done by the program.

- **The use of software is being considered and prototyped for other documents.**

Appendix B

28. In Data Summary Sheet #1, add a comment at the bottom of the table that describes the rationale for rejecting this run. This then becomes the record for the auditors.

- **A statement has been added as recommended.**

29. In Data Summary Sheet #4a, move the sentence to just above the table and modify it to read, "The coefficients in the table below are multipliers derived from the experimental design and should not be changed." It would also be helpful if the coefficient columns were "grayed" out.

- **This change has been made.**

30. In Data Summary Sheet #5a, a more explicit example is needed. It is too difficult to follow for the nonstatistician. Perhaps more explanation of where the numbers come from, put in straightforward language, would be helpful. The audience for this document is not as sophisticated as this document assumes.

- **The subcommittee believes that the document would become unwieldy if more explanation were provided. All calculations can be performed with a spreadsheet program without the need for worksheets. Future modifications of the document will either provide software or the steps needed to perform calculations in a spreadsheet program.**

Appendix C

31. In C2, a more detailed explanation of the multiple regression analysis, in "lay" terms, would be beneficial.

- **See response to Comment 31.**

Related NCCLS Publications*

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

EP6-P2 **Evaluation of the Linearity of Quantitative Analytical methods; Proposed Guideline (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-A **Interference Testing in Clinical Chemistry; Proposed Guideline (2002).** This document provides background information, guidance and experimental procedures for investigating, identifying, and characterizing the effects of interfering substances on clinical chemistry test results.

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

EP11-P **Uniformity of Claims for *In Vitro* Diagnostics Tests; Proposed Guideline (1996).** This document provides guidance to promote consistency in the content and interpretation of maximum performance claims for *in vitro* diagnostic testing systems.

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

NOTES

NOTES

NOTES

NOTES

NCCLS ▾ 940 West Valley Road ▾ Suite 1400 ▾ Wayne, PA 19087 ▾ USA ▾ PHONE 610.688.0100
FAX 610.688.0700 ▾ E-MAIL: exoffice@nccls.org ▾ WEBSITE: www.nccls.org ▾ ISBN 1-56238-482-1