EIGHTH EDITION

TIETZ FUNDAMENTALS OF

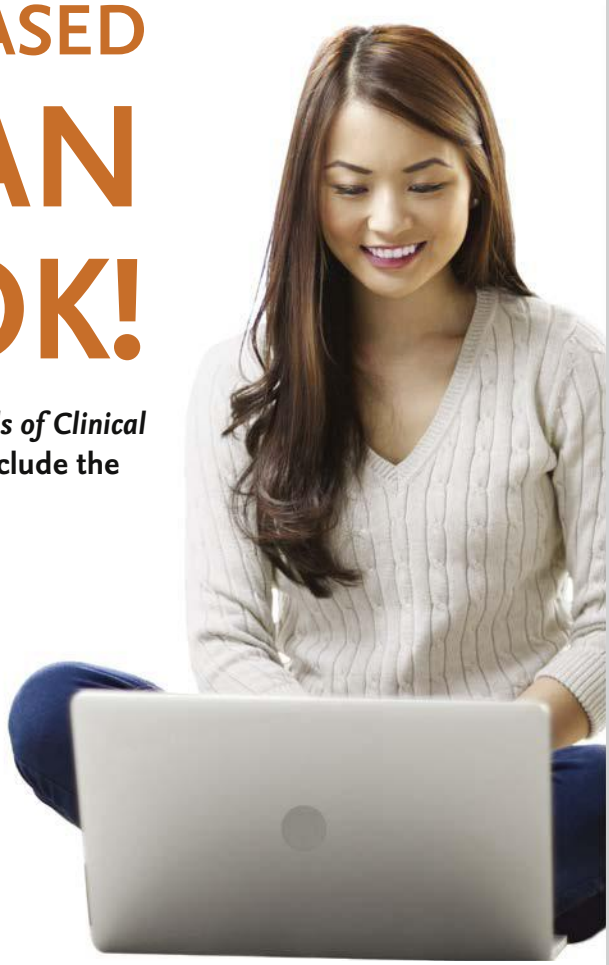
CLINICAL CHEMISTRY and MOLECULAR DIAGNOSTICS

Nader Rifai
Andrea Rita Horvath
Carl T. Wittwer

ELSEVIER

Evolve®

YOU'VE JUST PURCHASED MORE THAN A TEXTBOOK!



Evolve Student Resources for *Rifai: Tietz Fundamentals of Clinical Chemistry and Molecular Diagnostics, Eighth Edition*, include the following:

- Answers to Review Questions
- Case Studies
- The Coakley Collection of Clinical Cases
- Deacon's Challenge Biochemical Calculations Collection
- References

Activate the complete learning experience that comes with each NEW textbook purchase by registering with your scratch-off access code at

http://evolve.elsevier.com/Tietz/fundamentals/

If you rented or purchased a used book and the scratch-off code at right has already been revealed, the code may have been used and cannot be re-used for registration. To purchase a new code to access these valuable study resources, simply follow the link above.

Place Sticker Here

REGISTER TODAY!

ELSEVIER

You can now purchase Elsevier products on Evolve!
Go to evolve.elsevier.com/shop to search and browse for products.

2018v1.0

TIETZ FUNDAMENTALS OF

CLINICAL CHEMISTRY and MOLECULAR DIAGNOSTICS

This page intentionally left blank

EIGHTH EDITION

TIETZ FUNDAMENTALS OF CLINICAL CHEMISTRY and MOLECULAR DIAGNOSTICS

Nader Rifai, PhD
Professor of Pathology
Harvard Medical School
Louis Joseph Gay-Lussac Chair in Laboratory Medicine
Director of Clinical Chemistry
Boston Children's Hospital
Boston, Massachusetts

Andrea Rita Horvath, MD, PhD, FRCPath, FRCPA
Professor (Honorary), School of Public Health
University of Sydney
Professor (Conjoint), School of Medical Sciences
University of New South Wales
Clinical Director, SEALS Department of Clinical Chemistry and Endocrinology
New South Wales Health Pathology
Prince of Wales Hospital
Sydney, Australia

Carl T. Wittwer, MD, PhD
Professor of Pathology
University of Utah School of Medicine
Medical Director, Immunologic Flow Cytometry
ARUP Laboratories
Salt Lake City, Utah

ELSEVIER

ELSEVIER

3251 Riverport Lane
St. Louis, Missouri 63043

TIETZ FUNDAMENTALS OF CLINICAL CHEMISTRY AND
MOLECULAR DIAGNOSTICS, EIGHTH EDITION ISBN: 978-0-323-53044-6

International Standard Book Number: 978-0-323-53044-6

Director, Content Development: Laurie Gower
Executive Content Strategist: Kellie White
Publishing Services Manager: Catherine Jackson
Senior Project Manager: Amanda Mincher
Design Direction: Ryan Cook

Printed in the United States of America

Last digit is the print number: 9 8 7 6 5 4 3 2 1

ASSOCIATE EDITORS

This page intentionally left blank

CONTRIBUTORS

Aasne K. Aarsand, MD, PhD
Consultant Medical Biochemist
Norwegian Porphyria Centre and
Laboratory of Clinical Biochemistry
Haukeland University Hospital
Consultant Medical Biochemist
Norwegian Quality Improvement of
Laboratory Examinations
Haraldplass Deaconess Hospital
Bergen, Norway

Khosrow Adeli, PhD, FCACB, DABCC
Head and Professor, Clinical Biochemistry
The Hospital for Sick Children
University of Toronto
Toronto, Canada

Fred S. Apple, PhD
Medical Director of Clinical Laboratories
Hennepin County Medical Center
Professor of Laboratory Medicine and
Pathology
University of Minnesota
Minneapolis, Minnesota

Michael N. Badminton, BSc, MBChB, PhD, FRCPath
Clinical Senior Lecturer, School of
Medicine
Cardiff University
Consultant, Medical Biochemistry
University Hospital of Wales
Cardiff, Wales, United Kingdom

Tony Badrick, BAppSc, BSc, BA, MLitSt, MBA, PhD, PhD, FAACB, FAIMS, FRCPA (Hon), FACB, FFScRCPA
Associate Professor, Faculty of Health
Sciences and Medicine
Bond University, Robina
Queensland, Australia
Chief Executive Officer
Royal College of Pathologists of Australasia
Quality Assurance Programs
Sydney, Australia

Renze Bais, BSc(Hons), PhD, FFSc(RCPA), DipT
Director
RBaisconsulting.com
Queensland, Australia

Lindsay A.L. Bazydlo, PhD, DABCC
Assistant Professor of Pathology
University of Virginia School of Medicine
Director of Toxicology Laboratory
Associate Director of Clinical Chemistry
Laboratory
Co-Director of Coagulation Laboratory
University of Virginia Health System
Charlottesville, Virginia

Laura K. Bechtel, PhD, DABCC
CLIA Laboratory Director
Colorado Antiviral Pharmacology
Laboratory
University of Colorado Denver-Skaggs
School of Pharmacy and Pharmaceutical
Sciences
Laboratory Director
Cordant Health Solutions
Denver, Colorado

Marc Berg, MD, FAAP
Clinical Professor of Pediatrics
Division of Pediatric Critical Care Medicine
Stanford University, Lucile Packard
Children's Hospital
Palo Alto, California

Roger L. Bertholf, PhD
Professor of Clinical Pathology and
Laboratory Medicine
Weill Cornell Medicine
Medical Director of Clinical Chemistry
Department of Pathology and Genomic
Medicine
Houston Methodist Hospital
Houston, Texas

D. Hunter Best, PhD
Associate Professor of Pathology
University of Utah School of Medicine
Medical Director, Molecular Genetics &
Genomics
ARUP Laboratories
Salt Lake City, Utah

Ingvar Bjarnason, MD, MSc, FRCPath, FRCP, DSc
Professor of Digestive Disease
Consultant Physician and
Gastroenterologist
King's College Hospital
London, United Kingdom

Lee M. Blum, PhD
Assistant Laboratory Director/Toxicologist
Quality Assurance
NMS Labs, Inc.
Willow Grove, Pennsylvania

James C. Boyd, MD
Associate Professor (Emeritus)
Department of Pathology
University of Virginia Health System
Charlottesville, Virginia

Maria Domenica Cappellini, MD
Professor of Internal Medicine
Clinical Sciences and Community Health
University of Milan
Ca Granda Foundation-Policlinico Hospital
Milan, Italy

Ferruccio Ceriotti, MD
Laboratory Medicine Service
Ospedale San Raffaele
Milan, Italy

Mark A. Cervinski, PhD
Associate Professor, Pathology and
Laboratory Medicine
The Geisel School of Medicine at
Dartmouth
Hanover, New Hampshire
Director of Clinical Chemistry, Pathology
and Laboratory Medicine
Dartmouth-Hitchcock Medical Center
Lebanon, New Hampshire

Rossa W.K. Chiu, MBBS, PhD, FHKAM, FRCPA
Choh-Ming Li Professor of Chemical
Pathology
Department of Chemical Pathology
The Chinese University of Hong Kong
Shatin, New Territories
Hong Kong, China

Nigel J. Clarke, BSc (Hons), PhD
Vice President, Advanced Technology
Quest Diagnostics Nichols Institute
San Juan Capistrano, California

Timothy J. Cole, BSc(Hons), PhD
Associate Professor of Biochemistry and
Molecular Biology
Monash University
Clayton, Melbourne
Victoria, Australia

Mark Cooper, BMBCh, PhD, FRCP, FRACP
Professor of Medicine, Endocrinology
Concord Repatriation General Hospital
Professor of Medicine
University of Sydney
Sydney, Australia

Michael P. Delaney, BSc, MD, FRCP, LLM
Consultant Nephrologist, Renal Medicine
East Kent Hospitals University NHS Foundation Trust
Canterbury, Kent, United Kingdom

Sridevi Devaraj, PhD, DABCC, FAACC
Professor of Pathology and Immunology
Baylor College of Medicine
Medical Director, Clinical Chemistry and Point of Care
Texas Children's Hospital
Houston, Texas

Dennis J. Dietzen, PhD
Professor of Pathology and Immunology
Washington University School of Medicine
Medical Director of Laboratory Services
St. Louis Children's Hospital
St. Louis, Missouri

Paul D'Orazio, PhD
Director, Critical Care Analytical Research and Development
Instrumentation Laboratory
Bedford, Massachusetts

Graeme Eisenhofer, PhD
Professor and Chief, Division of Clinical Neurochemistry
Department of Medicine III and Institute of Clinical Chemistry and Laboratory Medicine
University Hospital Carl Gustav Carus Dresden at the Techniche Universität Dresden
Dresden, Germany

Christina Ellervik, MD, PhD, DMSci
Associate Director of Clinical Chemistry
Assistant Professor of Pathology
Harvard Medical School
Boston Children's Hospital
Boston, Massachusetts
Associate Professor of Clinical Chemistry
Department of Clinical Medicine
Faculty of Health and Medical Sciences
University of Copenhagen
Copenhagen, Denmark

Callum G. Fraser, BSc, PhD
Professor, Centre for Research into Cancer Prevention and Screening
University of Dundee
Dundee, United Kingdom

William D. Fraser, BSc (Hons), MB ChB, MD (Hons), FRCP, FRCPath, Eur Clin Chem
Professor of Medicine, Dean of Norwich Medical School
University of East Anglia
Honorary Consultant in Metabolic Medicine/Chemical Pathology
Department of Clinical Biochemistry
Norfolk and Norwich University Hospital Trust
Norwich, Norfolk, United Kingdom

Dr. Danielle B. Freedman, MB, BS, FRCPath, EuSpLM
Consultant Chemical Pathologist and Associate Physician in Clinical Endocrinology
Director of Pathology
Luton and Dunstable University Hospital
Luton, United Kingdom

Jonathan R. Genzen, MD, PhD
Associate Professor, Department of Pathology
University of Utah
Section Chief, Clinical Chemistry
ARUP Laboratories
Salt Lake City, Utah

Jens Peter Goetze, MD, DMSc
Professor, Chief Physician
Department of Clinical Biochemistry
University of Copenhagen, Rigshospitalet
Copenhagen, Denmark

Ann Gronowski, PhD
Professor of Pathology and Immunology
Professor of Obstetrics and Gynecology
Washington University School of Medicine
St. Louis, Missouri

David S. Hage, PhD
Professor, Department of Chemistry
University of Nebraska
Lincoln, Nebraska

David Halsall, PhD
Clinical Director, Blood Sciences
Addenbrooke's Hospital
Cambridge, United Kingdom

Doris M. Haverstick, PhD
Associate Professor of Pathology
University of Virginia
Charlottesville, Virginia

Charles D. Hawker, PhD, MBA, FACSc, FAACC
Adjunct Professor of Pathology (Retired), School of Medicine
University of Utah
Scientific Director
Automation and Special Projects (Retired)
ARUP Laboratories
Salt Lake City, Utah

Victoria Higgins, PhD Candidate
Laboratory Medicine and Pathobiology
University of Toronto
Toronto, Canada

Christopher P. Holstege, MD
Professor, Emergency Medicine and Pediatrics
University of Virginia School of Medicine
Chief, Division of Medical Toxicology
University of Virginia School of Medicine
Medical Director
Blue Ridge Poison Center
University of Virginia Health System
Charlottesville, Virginia

Dave S.B. Hoon, MSc, PhD
Professor, Director
Translational Molecular Medicine
Sequencing Center, John Wayne Cancer Institute
Saint John's Health Center
Providence Health System
Santa Monica, California

Gary L. Horowitz, MD
Director of Informatics
Department of Pathology and Laboratory Medicine
Professor of Pathology
Tufts Medical Center
Boston, Massachusetts

Andrea Rita Horvath, MD, PhD, FRCPath, FRCPA
Professor (Honorary), School of Public Health
University of Sydney
Professor (Conjoint), School of Medical Sciences
University of New South Wales
Clinical Director, SEALS Department of Clinical Chemistry and Endocrinology
New South Wales Health Pathology
Prince of Wales Hospital
Sydney, Australia

John Greg Howe, PhD
Associate Professor of Laboratory Medicine
Yale University School of Medicine
New Haven, Connecticut

Allan S. Jaffe, MD
Consultant in Cardiology and Laboratory Medicine
Professor of Medicine
Chair
CCLS Division of Laboratory Medicine and Pathology
Mayo Clinic and Medical School
Rochester, Minnesota

Graham Ross Dallas Jones, MBBS, BSc(med), DPhil, FRCPA, FAACB
Conjoint Associate Professor, Department of Chemical Pathology, SydPath
St. Vincent's Hospital
Faculty of Medicine
University of New South Wales
Sydney, Australia

Patricia M. Jones, PhD, DABCC, FAACC
Professor of Pathology
University of Texas Southwestern Medical Center
Clinical Director, Chemistry
Department of Pathology
Children's Medical Center
Dallas, Texas

Emily Jungheim, MD, MSCI
Associate Professor of Obstetrics and Gynecology
Washington University School of Medicine
St. Louis, Missouri

Todd William Kelley, MD, MS
Laboratory Director
Navican Genomics, Inc.
Salt Lake City, Utah

Mark D. Kellogg, PhD, MT(ASCP), DABCC, FAACC
Assistant Professor of Pathology
Harvard Medical School
Director of Quality Programs and Associate Director of Chemistry
Boston Children's Hospital, Department of Laboratory Medicine
Boston, Massachusetts

Larry J. Kricka, DPhil, FACB, CChem, FRSC, FRCPath
Professor Emeritus, Department of Pathology and Laboratory Medicine
University of Pennsylvania Medical Center
Philadelphia, Pennsylvania

Mark M. Kushnir, PhD
Senior Scientist
ARUP Institute for Clinical and Experimental Pathology
ARUP Laboratories
Adjunct Assistant Professor of Pathology
University of Utah School of Medicine
Salt Lake City, Utah

Edmund J. Lamb, BSc, MSc, PhD, FRCPath
Consultant Clinical Scientist, Laboratory Medicine
East Kent Hospitals University NHS Foundation Trust
Canterbury, Kent, United Kingdom

James P. Landers, BS, PhD
Professor of Chemistry
University of Virginia
Charlottesville, Virginia

Loralie Langman, PhD
Professor of Laboratory Medicine and Pathology
Mayo Clinic College of Medicine
Rochester, Minnesota

Evi Lianidou, PhD
Professor of Analytical Chemistry–Clinical Chemistry
National and Kapodistrian University of Athens
Athens, Greece

Kristian Linnet, MD, PhD
Professor of Forensic Chemistry
University of Copenhagen
Copenhagen, Denmark

Stanley Lo, PhD
Associate Professor of Pathology
Medical College of Wisconsin
Technical Director, Clinical Chemistry
Pathology and Laboratory Medicine
Children's Hospital of Wisconsin
Milwaukee, Wisconsin

Y.M. Dennis Lo, MA, DM, DPhil, FRCP (Lond. & Edin.), FRCPath, FRS
Li Ka Shing Professor of Medicine
Department of Chemical Pathology
The Chinese University of Hong Kong
Shatin, New Territories
Hong Kong, China

Nicola Longo, MD, PhD
Professor and Chief, Division of Medical Genetics
Department of Pediatrics
Adjunct Professor of Pathology
University of Utah School of Medicine
Co-Director, Biochemical Genetics and Newborn Screening Laboratories
ARUP Laboratories
Salt Lake City, Utah

G. Mike Makrigiorgos, PhD
Professor and Director of Medical Physics and Biophysics
Radiation Oncology, Dana Farber Cancer Institute
Harvard Medical School
Boston, Massachusetts

Elaine R. Mardis, PhD
Professor of Pediatrics, College of Medicine
Ohio State University
Co-Director, Institute for Genomic Medicine, Research Institute
Nationwide Children's Hospital
Columbus, Ohio

William J. Marshall, MA, PhD, MSc, MB, BS, FRCP, FRCPath, FRCPEdin, FRSC, FLS
Emeritus Reader in Clinical Biochemistry
King's College London
London, United Kingdom

Ann McCormack, MBBS Hons I, FRACP, PhD
Head
Hormones and Cancer Group
Garvan Institute of Medical Research
Endocrinologist
St Vincent's Hospital
Conjoint Senior Lecturer of Medicine
University of New South Wales
Sydney, Australia

Christopher McCudden, PhD, DABCC, FACB, FCACB
Clinical Biochemist
The Ottawa Hospital, Department of Pathology and Lab
University of Ottawa, School of Medicine
Ottawa, Ontario, Canada

Gwendolyn A. McMillin, PhD
Professor of Pathology
University of Utah
Medical Director, Toxicology and Pharmacogenomics
ARUP Laboratories
Salt Lake City, Utah

W. Greg Miller, PhD
Professor of Pathology
Virginia Commonwealth University
Richmond, Virginia

Michael C. Milone, MD, PhD
Associate Professor of Pathology and Laboratory Medicine
Perelman School of Medicine of the University of Pennsylvania
Philadelphia, Pennsylvania

Karel G.M. Moons, PhD
Professor of Clinical Epidemiology
Julius Centre for Health Sciences and Primary Care
University Medical Center Utrecht
Utrecht University
Utrecht, The Netherlands

Robert D. Nerenz, PhD, DABCC
Assistant Professor of Pathology and Laboratory Medicine
Dartmouth-Hitchcock Medical Center
Lebanon, New Hampshire

Frederick S. Nolte, PhD, D(ABMM), F(AAM)
Professor and Vice-Chair
Department of Pathology and Laboratory Medicine
Medical University of South Carolina
Charleston, South Carolina

Maurice O'Kane, BSc(Hons), MB ChB, MD, FRCPath, FRCPEdin
Consultant Chemical Pathologist
Clinical Chemistry Laboratory
Altnagelvin Hospital, Western Health and Social Care Trust
Londonderry, Northern Ireland, United Kingdom

Mauro Panteghini, MD
Professor of Clinical Biochemistry and Clinical Molecular Biology
Department of Biomedical and Clinical Sciences 'Luigi Sacco'
University of Milan Medical School
Director, Clinical Pathology Laboratory
ASST Fatebenefratelli-Sacco, Milan
Scientific Coordinator, Research Centre for Metrological Traceability in Laboratory Medicine (CIRME)
University of Milan
Milan, Italy

Jason Y. Park, MD, PhD
Director, Advanced Diagnostics Laboratory
Department of Pathology
Children's Medical Center Dallas
Associate Professor of Pathology
The Eugene McDermott Center for Human Growth and Development
University of Texas Southwestern Medical Center
Dallas, Texas

Marzia Pasquali, PhD, FACMG
Professor of Pathology
University of Utah School of Medicine
Medical Director, Biochemical Genetics and Newborn Screening
ARUP Laboratories
Salt Lake City, Utah

Jay L. Patel, MD, MBA
Associate Professor of Pathology
University of Utah School of Medicine
Medical Director
Hematopathology and Molecular Oncology
ARUP Laboratories
Salt Lake City, Utah

Victoria M. Pratt, PhD
Director, Pharmacogenomics and Molecular Genetics Laboratories
Medical and Molecular Genetics
Indiana University School of Medicine
Indianapolis, Indiana

Christopher P. Price, PhD, FRCPath, FRSC, FACB
Honorary Senior Fellow
Nuffield Department of Primary Care Health Sciences
University of Oxford
Oxford, United Kingdom

Alan T. Remaley, MD, PhD
Senior Investigator
National Institutes of Health
National Heart, Lung and Blood Institute
Bethesda, Maryland

Nader Rifai, PhD
Professor of Pathology
Harvard Medical School
Louis Joseph Gay-Lussac Chair in Laboratory Medicine
Director of Clinical Chemistry
Boston Children's Hospital
Boston, Massachusetts

Alan L. Rockwood, PhD, DABCC
President
Rockwood Scientific Consulting
Professor of Pathology (Retired)
University of Utah School of Medicine
Salt Lake City, Utah

Thomas Røraas, PhD
Mathematician
Norwegian Quality Improvement of Laboratory Examinations
Haraldsplass Deaconess Hospital
Bergen, Norway

Thomas Rosano, PhD
Professor of Pathology
Director of Laboratory Medicine
Albany Medical Center
Albany, New York

William Rosenberg, MA, MB, BS, DPhil, FRCP
Peter Scheuer Chair of Liver Diseases
Institute for Liver and Digestive Health
Division of Medicine
University College London
London, United Kingdom

David B. Sacks, MB, ChB, FRCPath
Adjunct Professor of Medicine
Georgetown University
Clinical Professor of Pathology
George Washington University
Washington, DC
Honorary Professor of Clinical Laboratory Sciences
University of Cape Town
Cape Town, South Africa
Senior Investigator
Chief, Clinical Chemistry
National Institutes of Health
Bethesda, Maryland

Sverre Sandberg, MD, PhD
Director, Norwegian Quality Improvement of Laboratory Examinations
Haraldsplass Deaconess Hospital
Professor, Institute of Public Health and Primary Health Care
University of Bergen
Director, Norwegian Porphyria Centre
Haukeland University Hospital
Bergen, Norway

Mitchell G. Scott, PhD
Professor of Pathology and Immunology
Division of Laboratory and Genomic Medicine
Co-Medical Director, Clinical Chemistry
Washington University School of Medicine
St. Louis, Missouri

Leslie M. Shaw, BS, PhD
Professor of Pathology and Laboratory Medicine
Director, Toxicology Laboratory
Director, Biomarker Research Laboratory
Pathology and Laboratory Medicine
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania

Roy A. Sherwood, BSc, MSc, DPhil
Professor of Clinical Biochemistry
King's College London
London, United Kingdom

Ana-Maria Simundic, PhD
Professor, Specialist in Laboratory Medicine
Department of Medical Laboratory Diagnostics
Clinical Hospital Sveti Duh
Zagreb, Croatia

Ravinder Sodi, PhD, CSci, EuSpLM, FRCPath
Consultant Clinical Biochemist
Department of Blood Sciences
Royal Lancaster Infirmary and Furness General Hospital
University Hospitals of Morecambe Bay NHS Foundation Trust
Honorary Lecturer
Lancaster Medical School
University of Lancaster
Lancaster, United Kingdom

Andrew St. John, PhD
Principal Consultant
ARC Consulting
Perth, Australia

Molly J. Stout, MD, MSCI
Assistant Professor of Obstetrics and Gynecology
Washington University School of Medicine
St. Louis, Missouri

Frederick G. Strathmann, PhD, MBA
Vice President of Quality Assurance
Director of New Technology and Innovation
Assistant Laboratory Director
NMS Labs
Willow Grove, Pennsylvania

Catharine Sturgeon, BSc, PhD, FRCPath
Honorary Lecturer
University of Edinburgh
Consultant Clinical Scientist
Department of Laboratory Medicine
Royal Infirmary of Edinburgh
Edinburgh, United Kingdom

Dorine W. Swinkels, MD, PhD
Professor of Experimental Clinical Chemistry
Department of Laboratory Medicine
Translational Metabolic Laboratory
Radboudumc, Nijmegen, The Netherlands

Sudeep Tanwar, MD, MRCP(UK)
Consultant Gastroenterologist and Hepatologist
Barts Health NHS Trust
London, United Kingdom

Andrew Taylor, PhD, FRSC, FRCPath
Consultant Clinical Biochemist
Royal Surrey County Hospital
Visiting Reader
University of Surrey
Guildford, Surrey, United Kingdom

Mia Wadelius, MD, PhD
Professor of Clinical Pharmacogenetics
Department of Medical Sciences, Clinical Pharmacology
Uppsala University
Uppsala, Sweden

Natalie E. Walsham, MBiochem, MSc, DipRCPath
Lead Clinical Scientist, Department of Clinical Biochemistry
University Hospital Lewisham
London, United Kingdom

Sharon D. Whatley, MSc, PhD, FRCPath
Clinical Scientist
All Wales Genetics Laboratory
Institute of Medical Genetics
University Hospital of Wales
Cardiff, Wales, United Kingdom

Ronald J. Whitley, PhD, DABCC, FAACC
Professor Emeritus, Department of Pathology and Laboratory Medicine
University of Kentucky
Lexington, Kentucky

William E. Winter, MD, DABCC, FAACC, FCAP
Professor of Pathology, Immunology, and Laboratory Medicine, Pediatrics, and Molecular Genetics and Microbiology
University of Florida
Gainesville, Florida

Carl T. Wittwer, MD, PhD
Professor of Pathology
University of Utah School of Medicine
Medical Director, Immunologic Flow Cytometry
ARUP Laboratories
Salt Lake City, Utah

Melanie L. Yarbrough, PhD
Clinical Chemistry Fellow
Department of Pathology and Immunology
Washington University School of Medicine
St. Louis, Missouri

CONTRIBUTORS TO EVOLVE

Mark D. Kellogg, PhD, MT(ASCP), DABCC, FAACC
Assistant Professor of Pathology
Harvard Medical School
Director of Quality Programs and Associate Director of Chemistry
Boston Children's Hospital, Department of Laboratory Medicine
Boston, Massachusetts

Roy W.A. Peake, PhD, FRCPath D(ABCC), D(ABMGG)
Associate Director of Clinical Chemistry
Boston Children's Hospital
Instructor
Harvard Medical School
Boston, Massachusetts

Wenfang S. Wu, PhD
Clinical Chemistry Postdoctoral Research Fellow
Harvard Medical School
Boston Children's Hospital, Department of Laboratory Medicine
Boston, Massachusetts

REVIEWERS

Daniel Bankson, SM, PhD, MBA
Associate Professor
Laboratory Medicine
University of Washington
Seattle, Washington

Janelle M. Chiasera, PhD
Professor and Chair
Senior Executive Associate Dean
School of Health Professions/Clinical and Diagnostic Sciences
The University of Alabama at Birmingham
Birmingham, Alabama

Katherine Davis, MS, MT(ASCP), CLS
Assistant Professor
Department of Clinical Laboratory Science
Senior Assessment Specialist
Academic Dean's Office
Loma Linda University
Loma Linda, California

Paula Deming, PhD
Associate Professor and Chair
Medical Laboratory and Radiation Sciences
Endowed Professor of Health Sciences
University of Vermont
Burlington, Vermont

Karen M. Escolas, EdD, MT(ASCP)
Chair
Department of Medical Laboratory Technology
Farmingdale State College
Farmingdale, New York

Vicky S. Freeman, PhD
Chair and Professor
Clinical Laboratory Sciences
University of Texas Medical Branch
Galveston, Texas

Shawn M. Froelich, MS, MLS(ASCP)^{CM}
Clinical Coordinator
UnityPoint Health
Assistant Professor
Medical Laboratory Science
Allen College
Waterloo, Iowa

Neil Selwyn Harris, MBChB, MD
Clinical Associate Professor
Department of Pathology, Immunology and Laboratory Medicine
University of Florida, College of Medicine
Gainesville, Florida

Kathleen McLoughlin Kenwright, EdD, MLS(ASCP)MB^{CM}
Chair and Associate Professor
Clinical Laboratory Science
University of Tennessee Health Science Center
Memphis, Tennessee

Nadine M. Lerret, PhD, MLS(ASCP)^{CM}
Medical Laboratory Science
Rush University
Chicago, Illinois

Mary Ann McLane, BS, MS, PhD
Professor
Medical Laboratory Sciences
University of Delaware
Newark, Delaware

Nicholas M. Moore, MS, MLS(ASCP)^{CM}
Assistant Professor and Assistant Director
Departments of Pathology and Medical Laboratory Science
Rush University Medical Center
Chicago, Illinois

Michelle Moy, MAdEd, MT(ASCP)SC
Assistant Professor and Program Director Biomedical Sciences
Department of Biological and Health Sciences
Madonna University
Livonia, Michigan

Kathleen Park, MA, MT(ASCP)
Assistant Department Chair
Medical Laboratory Technology
Austin Community College
Austin, Texas

Gerald DeWayne Redwine, PhD, MEd, MT(ASCP)
Associate Professor
Clinical Laboratory Science
Texas State University
San Marcos, Texas

Brady G. Rogers, MSBMS, MT
Medical Laboratory Technician Program Chair
Medical Laboratory Technology
Fortis College
Landover, Maryland

Robin Thompson-McAvoy, MLT, BSc, BEd-Adult
Coordinator of Medical Laboratory Science
Professor of Clinical Chemistry
St. Lawrence College
Kingston, Ontario

PREFACE

We are pleased to introduce the eighth edition of *Tietz Fundamentals of Clinical Chemistry and Molecular Diagnostics*. We built on the excellent work of our predecessors and used electronic tools to produce a state-of-the-art product for students, trainees, and practicing clinical laboratory scientists.

Chapters were extensively updated, and more than 60 new authors were recruited to present the most current and relevant information. We aimed to harmonize the presentation of information among chapters while retaining the personality and unique style of each author, hoping for a readable, educational text.

Unlike most other textbooks, all chapters in this edition were reviewed by three individuals: a reviewer, an associate editor, and a senior editor. We believe that these efforts have led to a better product. In addition, we made a concerted effort to create an *international* rather than an *American* product to reflect different practices from around the world; for example, all measurements are presented both in traditional and SI units.

In addition to the print format of *Fundamentals*, a wealth of supplementary educational materials including clinical case studies, biochemical calculations, multiple-choice questions, and references are available on the Elsevier Evolve platform for an enhanced learning experience.

This project has been a true group effort and represents the collective intellect, knowledge, and experience of approximately 120 leaders in laboratory medicine from 13 countries. We are in debt not only to the authors, reviewers, and editors of the chapters but also to the contributors of the supplementary materials that greatly enrich the product. We are grateful to Elsevier, and particularly to Laurie Gower, for supporting us throughout this project.

We sincerely hope that this product will be a valuable educational and reference resource for the clinical laboratory scientists' community worldwide.

Nader Rifai
Andrea Rita Horvath
Carl T. Wittwer

This page intentionally left blank

CONTENTS

PART V Molecular Diagnostics

1

Clinical Chemistry and Molecular Diagnostics

*

1. Define the following terms:
   - Ethics
   - Laboratory medicine
   - Molecular diagnostics
2. List and explain the reasons for performing a laboratory test.
3. Describe the field of laboratory medicine, including subdisciplines, information handling, and ethical issues.
4. Describe the role of the clinical chemist.
5. Describe the possible career paths for the clinical chemist.
6. State the applications of molecular diagnostics in laboratory medicine.
7. List and explain five ethical issues that confront laboratorians; describe the critical importance of maintaining confidentiality in the laboratory.
8. Evaluate a possible confidentiality or conflict of interest issue and determine whether it is an ethics violation.
9. State the roles of authors, editors, reviewers, and publishers in providing high quality scientific publications.

Ethics Rules or standards governing the conduct of an individual or the members of a profession.

Laboratory medicine A component of laboratory science that is involved in the selection, provision, and interpretation of diagnostic testing of individual specimens.

Laboratory testing A process conducted in a clinical laboratory to rule in or rule out a diagnosis, to select and monitor disease treatment, to provide a prognosis, to screen for a disease, or to determine the severity of and monitor a physiological disturbance.

Molecular diagnostics Use of molecular biology techniques to predict, prevent, diagnose, and monitor disease, including the selection and optimization of therapies.

The disciplines of *clinical chemistry* and *molecular diagnostics* elicit different images. For clinical chemistry, one thinks of pH measurements or large chemistry analyzers, whereas molecular diagnostics conjures up the human genome project, companion diagnostics, and personalized and precision medicine. Although clinical chemistry is at the core of laboratory medicine, molecular diagnostics is a more recent but explosive upstart. Clinical chemistry excels in random access testing, but molecular diagnostics has evolved massively parallel methods. On the surface, these disciplines appear clearly different.

However, consider the meaning behind the words that compose "clinical chemistry" and "molecular diagnostics." Chemistry by its very nature is molecular, and the study of molecules is chemistry. There is no difference here. Perhaps the "molecular" in "molecular diagnostics" suggests complex polymers with meaningful sequence, excluding simpler chemicals. DNA and RNA sequences largely define life, and powerful technologies for nucleic acids now eclipse those for other complex polymers such as proteins and carbohydrates. In common parlance, molecular diagnostics is dominated by nucleic acids. The words "clinical" and "diagnostics" are also similar, connecting both fields to human disease. "Clinical" is more generic than "diagnostics," but again in common use, "molecular diagnostics" includes not only diagnostics but prognosis and genetic predisposition as well. In each

* The authors gratefully acknowledge the contributions by David E. Bruns, François A. Rousseau, and Carl A. Burtis on which portions of this chapter are based.

two-word combination, the sum is greater than its parts, with combined meanings evolving to fit needs and interest. We believe that molecular diagnostics is best viewed as a subset of clinical chemistry.

According to the definition of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), "Clinical Chemistry is the largest subdiscipline of Laboratory Medicine which is a multidisciplinary medical and scientific specialty with several interacting subdisciplines, such as hematology, immunology, clinical biochemistry, and others. Through these activities clinical chemists influence the practice of medicine for the benefit of the public."

Clinical laboratories provide in vitro testing of chemical, biochemical, and genetic markers in various fluids or tissues of the human body to screen for a disease, confirm or exclude a diagnosis, help to select or monitor a treatment, or assess prognosis. Laboratory testing impacts health care delivery to virtually every patient.

LOOKING BACK

The examination of body fluids for the diagnosis of disease is certainly not a modern concept. The Greeks noticed before 400 BC that ants are attracted to "sweet urine." However, laboratory testing was not always appreciated by clinicians; the famous Dublin physician Robert James Graves (1796–1853) once remarked, "Few and scanty, indeed, are the rays of light which chemistry has flung on the vital mysteries," and the pioneer Max Josef von Pettenkofer (1818–1901) stated that clinicians use their chemistry laboratory services only when needed for "luxurious embellishment for a clinical lecture." Such views have changed throughout the years, and laboratory testing has proven to be a useful tool to clinicians who have grown to depend and rely on laboratory testing in the routine management of their patients Box 1.1.

Although it may be difficult to pinpoint the exact date when the concept of the clinical laboratory was born, all indications point to the mid-19th century. One such indication is an article titled "Hospital Construction" by Francis H. Brown that was published in the *Boston Medical and Surgical Journal*, the precursor of the *New England Journal of Medicine*, in 1861. Dr. Brown stated, "[Every hospital should have] a small room at the end of the ward to serve as a general laboratory … necessary small cooking might be accomplished here; dishes and other articles washed, etc.; and it would serve as a general store-room for brooms, pails, and other articles." Although Baron Justus von Liebig (1803–73) boasted that his clinical laboratory performed more than 400 tests per annum, the average mid- to large-sized laboratory nowadays performs several million tests yearly. The term *clinical chemistry* was purportedly coined by Charles Henry Ralfe (1842–96) of London Hospital when he used it as the title of his 1883 treatise. The first laboratory attached to a hospital was established in 1886 in Munich, Germany, by Hugo Wilhelm von Ziemssen. In the United States the first clinical laboratory was The William Pepper Laboratory of Clinical Medicine, established in 1895 at the University of Pennsylvania in Philadelphia.

Molecular diagnostics has more recent origins. "Molecular diagnosis" was first mentioned in 1968 as the title of a *New England Journal of Medicine* editorial, commenting on a new inborn error of metabolism that overproduced oxalic acid, resulting in kidney stones. "Molecular" referred to an enzymatic pathway and the substrates, not nucleic acid variants. Twenty years later, additional articles describing "molecular diagnostics" began to appear. In 1986, molecular diagnostics was defined as, "… the detection and quantification of specific genes by nucleic acid hybridization procedures," exemplified by speciation of plant nematodes. In 1987, molecular diagnostics was used to describe mapping of antigenic substances by affinity chromatography using immobilized antibodies. In 1988 the term was used to describe methods for detecting gene amplification and rearrangement using Southern blotting. With the advent of polymerase chain reaction (PCR), the term "molecular diagnostics" became more common, its use doubling in the medical literature every 6 to 7 years. By 1997, commercial real-time PCR instruments solidified "molecular diagnostics" as a branch of clinical chemistry and laboratory medicine.

BOX 1.1 Uses of Testing in the Clinical Laboratory

- Confirming a clinical suspicion (which could include making a diagnosis)
- Excluding a diagnosis
- Assisting in selection, optimization, and monitoring of treatment
- Providing a prognosis
- Screening for disease in the absence of clinical signs or symptoms
- Establishing and monitoring the severity of a physiological disturbance

EXPANDING BOUNDARIES DEFINED BY TECHNOLOGY

Unlike other specialties in laboratory medicine, clinical chemistry is very much influenced and shaped by technology. No discipline in laboratory medicine uses more technologies than clinical chemistry. Technologies that evolved over time not only changed practice but also remodeled the boundaries of the traditional clinical chemistry laboratory. For example, with the emergence of immunochemical techniques in the 1970s, the US Food and Drug Administration approved many tests for the measurement of proteins, small molecule hormones, and drugs, a development that profoundly changed clinical chemistry and its armamentarium of testing. Integrated automated platforms later enabled the measurement of hormones and therapeutic drugs by immunoassays simultaneously with electrolytes, glucose, and other general chemistry tests, thus subsuming the "endocrine lab" and the "drug lab."

Serologic tests for hepatitis and human immunodeficiency virus (HIV) and tests for autoimmune diseases also moved from their traditional home in microbiology and immunology to chemistry analyzers. Immunoglobulin analysis followed a similar path. The typical clinical chemistry laboratory includes testing for general chemistries, specific proteins and immunoglobulins, therapeutic and abused drugs, blood gases, hormones, biogenic amines, porphyrins, vitamins, and trace elements. Testing for inborn errors of metabolism (such as the measurements of amino acids and organic acids), measurements of coagulation factors, general hematologic testing, and serologic assays can belong either to the clinical chemistry laboratory or to another subspecialty, depending on the institution.

Clinical chemists have embraced technology over the years and used it effectively to derive answers to clinical questions. In modern clinical chemistry laboratories, technologies include spectrophotometry, atomic absorption, flame emission photometry, nephelometry, electrochemical and optical sensor technologies, electrophoresis, and chromatography. The influence of automation, information technology, and miniaturization is evident in current clinical chemistry laboratories. Mass spectrometry, once thought of as a research tool, is playing an ever-growing role in clinical chemistry for the measurement of both small molecules and peptides, and more recently proteins. Point-of-care testing is a disruptive innovation that decentralizes laboratory testing and presents the clinical chemist with many challenges and opportunities.

Molecular diagnostics has forever changed virology and microbiology, introducing faster and more sensitive methods based on nucleic acid amplification rather than microbial replication. Nanotechnology, microfluidics, electrical impedance, reflectance spectroscopy, and time-resolved fluorescence are only a few of the technologies used in point-of-care testing for proteins, drugs, DNA, and analysis of metabolites in small samples of whole blood. Molecular diagnostics in particular impacts diverse specialties, including infectious disease, genetics, and oncology, providing new tools for study at a molecular detail never before considered. In summary, the boundaries of clinical chemistry expand with technology, making the profession vibrant, interesting, and ever evolving.

The scope of the profession is constantly changing for the very same reasons. Scientific and technological developments, medical needs, patient demands, and economic pressures bring various disciplines of medicine closer together, and this integration results in more effective health care. For example, companion diagnostics, which help predict therapeutic responses and individualize patient treatment options, bring together pharmacy and medical laboratories. Point-of-care testing in real time with medical intervention breaks the walls of laboratories to bring the profession closer to clinicians and patients. New disruptive technologies (e.g., "lab on a chip," nanotechnology, home monitoring) as well as movement toward patient empowerment and direct-to-consumer testing bring laboratory testing closer to patients. All of these developments present special challenges to the future generations of laboratory professionals both in terms of how they should be trained and how they will practice.

Technology alone is not the answer to more effective and cost-effective clinical practice. The laboratory data obtained must be meaningful and support clinical management decisions. The generation of more data does not necessarily lead to better patient management and outcomes. In the 1960s and 1970s, with the advent of automated clinical analyzers, laboratories reported chemistry panels of 10 to 20 results. More recently, dense data from expression arrays, genome-wide association studies, epigenomics, and microRNA analyses excel in discovery research, but translation to clinical practice has been slower than anticipated. The promise of greater clinical significance with larger data sets seems intuitive, but history suggests caution. Clinical chemists in this world of "big data" translate high-quality measurement *data* into clinically relevant *information*. This information—when integrated with clinical history and presentation, clinical signs, and an understanding of pathophysiology—becomes *knowledge*. Knowledge, in the context of the experience and judgment of the clinician, is converted to *wisdom* that translates to clinical action for improved patient outcomes.

HOW IS CLINICAL CHEMISTRY PRACTICED?

Although the majority of clinical chemists choose a career in a clinical laboratory environment, many work in the in vitro diagnostics (IVD) and pharmaceutical industries. Clinical chemists, by virtue of their training, are translational researchers who are capable of developing, evaluating, and validating biochemical and genetic assays for clinical use; they develop skills that are essential for new biomarker assays, reagent kits, and companion diagnostics. Clinical chemists also provide interfaces between researchers, clinicians, the clinical laboratory, and the IVD industry to help translate biomarker research into clinically meaningful decisions and actions.

Clinical chemists practicing in the IVD or the pharmaceutical industry may not need to routinely interact with clinicians or interpret laboratory results, but they understand and appreciate the clinical utility and relevance of the assays and companion diagnostics they are developing and thus contribute more effectively to the development of diagnostics that improve health. The daily practice of the profession has changed over time. In the 1960s and 1970s, clinical chemists developed laboratory tests. At present, de novo assay development is still active only in certain areas such as chromatography, mass spectrometry, and molecular diagnostics.

However, as the profession matured and the instrumentation changed from open systems to "black boxes," the traditional analytical focus of the profession has significantly diminished. Clinical chemists are now more active in the preanalytical and postanalytical phases of testing and in establishing processes such as how best to select the right test for the right patient and to communicate test results to clinicians in a medically meaningful way, how to build laboratory processes that reduce error, and how to continuously improve the quality of laboratory practices (Box 1.2).

BOX 1.2 Functions of the Laboratory Professional

Develop and validate de novo laboratory tests to meet clinical needs.
Evaluate and characterize the analytical and clinical performance of laboratory tests.
Present laboratory results to clinicians in an effective manner.
Provide education and advice on the selection and interpretation of laboratory tests as part of the clinical team.
Determine the cost effectiveness and intrinsic value of laboratory tests.
Participate in the development of clinical testing algorithms and clinical practice guidelines.
Ensure compliance with regulatory requirements.
Participate in quality assurance and improvement of the laboratory service.
Teach and train future generations of laboratory specialists.
Participate in basic or clinical research.

In the current health care environment, there is increasing emphasis on clinical impact and cost effectiveness. Laboratories are expected to demonstrate evidence of improved measurable clinical outcomes and the usefulness and added value of tests to clinical decision making. Proving the fact that laboratory testing contributes to improved patient outcomes is challenging because the relationship between testing and clinical outcomes is mostly indirect. Nevertheless, clinical chemists should move away from being just providers of high-quality data. Transforming laboratory data to information and knowledge requires more skills in information and information management technology, evidence-based medicine, epidemiology, data mining, and translational research. It also requires a shift of thinking from essentialism to consequentialism and from technology-driven to customer-focused and patient-centered laboratory medicine.

To summarize, today's clinical chemists are laboratory professionals who are trained in pathophysiology and technology. The execution of their daily duties, which are more clinically or technology oriented, is influenced by their training (such as MD vs. PhD), interests, institutional needs, and the country where they practice. Clearly the practice of our profession has evolved over the past half a century, and there are even more challenges on the horizon that will expand and change its scope and role and enhance its diversity.

GUIDING PRINCIPLES OF PRACTICING THE PROFESSION

As in other branches of medicine, practitioners in the clinical laboratory are faced with ethical issues, often on a daily basis; examples are listed in Box 1.3.

CONFIDENTIALITY OF GENETIC INFORMATION

Confidentiality of genetic information has been prominent in the news in the first and second decades of this millennium.

BOX 1.3 Ethical Issues in Clinical Chemistry and Molecular Diagnostics

Confidentiality of genetic information
Confidentiality of patient medical information
Allocation of resources
Codes of conduct
Publishing issues
Conflicts of interest

Legislation was considered necessary to prevent denial of health insurance or employment to people found by DNA testing to be at risk of disease. Less appreciated is the fact that the issue of confidentiality of clinical laboratory data predated DNA testing. In fact, many non-DNA tests, old and new, also carry information about risks of illness and death. Clinical laboratory professionals have long been responsible for maintaining the confidentiality of all laboratory results, a situation made even more critical with the advent of increasingly powerful genetic testing.

CONFIDENTIALITY OF PATIENT MEDICAL INFORMATION

New test development requires the use of patient samples and may involve the use of patient medical information. Ethical judgments are required regarding the type of informed consent that is needed from patients for use of their samples and clinical information. Clinical laboratory physicians and scientists often serve on institutional review boards that examine proposed research on human subjects. In these discussions, ethical concepts such as clinical equipoise—which refers to the genuine uncertainty in the expert medical community over whether a particular treatment will be beneficial—and confidentiality are central to decisions.

ALLOCATION OF RESOURCES

Because resources are finite, laboratory professionals must make ethically responsible decisions about allocation of resources. There is often a trade-off between cost and quality. What is best for patients generally? How can the most good be done with the available resources? For laboratorians in business, creative accounting may tarnish the profession if patient care is not kept paramount.

CODES OF CONDUCT

Most professional organizations publish a code of conduct that requires adherence by their members. For example, the American Association for Clinical Chemistry (AACC) has published ethical guidelines that require AACC members to endorse principles of ethical conduct in their professional activities, including (1) selection and performance of clinical procedures, (2) research and development, (3) teaching, (4) management, (5) administration, and (6) other forms of professional service.

PUBLISHING ISSUES

Publication of documents having high scientific integrity depends on editors, authors, and reviewers all working in concert in an environment governed by high ethical standards.

Editors are responsible for the overall process, including identifying reviewers, evaluating the reviews and the authors' response to them, and making the final decision of whether to accept or reject a manuscript. Editors are also responsible for establishing policies and procedures to ensure consistency in the editorial process. Finally, the editor-in-chief is responsible for developing a conflict of interest policy and monitoring it among his or her editors. Publishers, being commercial or scientific societies, should monitor any conflicts of interest of the editor-in-chief.

Authors are responsible for honest and complete reporting of original data produced in ethically conducted research studies. Practices such as fraud, plagiarism (verbatim, mosaic), and falsification or fabrication of data are unacceptable. The International Committee of Medical Journal Editors and the Committee on Publication Ethics have published policies that address such behavior. Other practices to be avoided include duplicate publication, redundant publication, and inappropriate authorship credit. In addition, ethical policies require that factors that might influence the interpretation of study findings must be revealed, such as (1) the role of the commercial sponsor in the design and conduct of the study, (2) interpretation of results, and (3) preparation of the manuscript. Additional undesirable and harmful practices are publication bias and selective reporting, in which only studies with positive findings are reported and authors use "data dredging" and meaningless subanalyses to find a positive association rather than reporting the original hypothesis that was negative. These practices inflate the actual value of observations or utility of markers and diminish the quality of meta-analyses. As a result, a comprehensive registry of diagnostic and prognostic studies, similar to the registry of clinical trials, has been advocated.

Reviewers must provide a timely, fair, and impartial assessment of manuscripts. They must maintain confidentiality and never contact the authors until after the publication of the report. Finally, reviewers must excuse themselves from the review process if they perceive a conflict of interest.

Most journals now require authors to complete conflict of interest forms and delineate each author's contribution. Some journals, including *Clinical Chemistry,* publish this information along with the article for enhanced transparency.

CONFLICTS OF INTEREST

The interrelationships between practitioners in the medical field and commercial suppliers of drugs, devices, and equipment can be positive or negative. Concerns led the National Institutes of Health in 1995 to require official institutional review of financial disclosure by researchers and management in situations when disclosure indicates potential or actual conflicts of interest. In 2009 the Institute of Medicine issued a report that questioned inappropriate relationships between pharmaceutical device companies and physicians and other health care professionals. Similarly, the relationship between clinical laboratory professionals and manufacturers and providers of diagnostic equipment and supplies can be scrutinized.

As a consequence of these concerns and as a result of the enactment of various laws designed to prevent fraud, abuse, and waste in Medicare, Medicaid, and other health care reimbursement programs, professional organizations that represent manufacturers of IVD and other device and health care companies have published codes of ethics. For example, the Advanced Medical Technology Association has published a revised code of ethics that became effective on July 1, 2009. Topics discussed in this revised code include gifts and entertainment, consulting arrangements and royalties, reimbursement for testing, and education. Similarly, the European Diagnostic Manufacturers Association has published a code of ethics, including issues of member-sponsored product training and education, support for third-party educational conferences, sales and promotional meetings, consultants, gifts, provision of reimbursements, and donations for charitable and philanthropic purposes. Both documents address demands from regulators while nurturing the unique role that laboratory and other health care professionals play in developing and refining new technology.

WHAT IS IN THIS TEXTBOOK?

In this textbook, we have assembled what is essential to effectively practice clinical chemistry and molecular diagnostics. We begin with introductory chapters that describe the basics of laboratory medicine, including statistics, sample handling, preanalytical processes, reference intervals, and quality control. This is followed by a section on analytical techniques and applications, describing the main methods used in clinical chemistry, including immunoassays, mass spectrometry, and point-of-care testing. Next, all the major analytes are discussed, including enzymes, tumor markers, therapeutic drugs, and toxicology, among many others. This is followed by a section on pathophysiology that covers disease states and malfunction of different organ systems that correlate with abnormal laboratory findings. Finally, our last section is dedicated to molecular diagnostics, perhaps the fastest growing field in clinical chemistry. An appendix tabulates reference intervals for the clinical laboratory. The online version includes clinical cases, podcasts, and biochemical calculations. Our aim is to provide current scientific and practical knowledge to support laboratory professionals with knowledge resources that interface between science and technology on the one hand and the clinician and the patient on the other.

POINTS TO REMEMBER

Clinical chemistry is the largest subdiscipline of laboratory medicine, and molecular diagnostics is a subset of clinical chemistry.
Clinical chemistry is a profession that has been shaped and defined by technology.
The role of clinical chemists evolved over time from analytically and technology focused to customer and patient centered.
Clinical chemists are translational researchers who convert laboratory data to clinical knowledge.
Career paths of clinical chemists are heterogeneous and include work in clinical laboratories and in vitro diagnostics and pharmaceutical industries.
Clinical chemists must adhere to guiding principles of practicing the profession, which include maintaining confidentiality of genetic and medical information, using resources appropriately, abiding by codes of conduct, following ethical publishing rules, and managing and disclosing conflict of interest.

REVIEW QUESTIONS

1. The clinical laboratory discipline that is used most often to assess inherited disease through study of the constitutive genome is:
   a. transfusion services.
   b. clinical chemistry.
   c. molecular diagnostics.
   d. hematology.
2. When a practitioner in clinical chemistry has an inappropriate personal relationship with a commercial supplier of chemistry analyzers, there may be a potential issue with:
   a. publication ethics.
   b. confidentiality.
   c. selection of treatment.
   d. conflict of interest.
3. "Molecular testing" involves the clinical analysis of:
   a. atoms and molecules.
   b. nucleic acids.
   c. cellular components of blood.
   d. the physical structure of compounds.
4. Which one of the following is not considered an ethical issue facing a clinical laboratorian?
   a. Allocation of resources
   b. Conflicts of interest
   c. Discussion of one's salary
   d. Maintenance of confidentiality

SUGGESTED READINGS

Altman, D. G. (2014). The time has come to register diagnostic and prognostic research. *Clinical Chemistry*, *60*, 580–582.

Annesley, T. M., Boyd, J. C., & Rifai, N. (2009). Publication ethics: clinical chemistry editorial standards. *Clinical Chemistry*, *55*, 1–4.

Hallworth, M. J., Epner, P. L., Ebert, C., et al. (2015). Current evidence and future perspectives on the effective practice of patient-centered laboratory medicine. *Clinical Chemistry*, *61*(4), 589–599.

McMurray, J., Zerah, S., Hallworth, M., et al. (2010). The European Register of Specialists in Clinical Chemistry and Laboratory Medicine: guide to the Register, version 3-2010. *Clinical Chemistry and Laboratory Medicine*, *48*, 999–1008.

Rifai, N., Annesley, T., & Boyd, J. (2010). International year of chemistry 2011: *Clinical Chemistry* celebrates. *Clinical Chemistry*, *56*, 1783–1785.

2

Analytical and Clinical Evaluation of Methods

1. Define the following:
   - Analytical measurement range
   - Analytical specificity
   - Bias
   - Coefficient of variation
   - Correlation coefficient
   - Difference curve
   - Error model
   - Frequency distribution
   - Gaussian probability distribution
   - Limit of detection
   - Linearity
   - Mean
   - Median
   - Population
   - Precision
   - Random error
   - Random sample
   - Regression analysis
   - Sample
   - Standard deviation
   - Systematic error
   - Student *t* distribution
   - Trueness
   - Uncertainty
2. List and describe three criteria that must be considerations in laboratory method selection, including the specific parameters involved in each criterion.
3. Compare population and sample mean, population parameter and sample statistic, and population standard deviation and sample standard deviation, including a description of each, symbols used to express these, how they are calculated, and the information they provide.
4. State the connection of the following concepts to analytical methods:
   - Accuracy
   - Analytical sensitivity
   - Analytical specificity
   - Calibration
   - Limit of detection
   - Linearity
   - Precision
   - Repeatability
   - Reproducibility
5. List two common approaches used to objectively analyze data in a methods comparison study.
6. Describe the components of a difference plot, including the plot's use in method comparison and how the plot is interpreted.
7. Discuss assessment of error in an objective analysis of data in method comparison, including how error occurrence relates to an assay's performance characteristics, the difference between random and systematic error, what causes error, and how error is evaluated in a difference plot.
8. For the following types of analyses, list the components of the analysis, its application in method comparison, how it is computed, how outliers affect it, and how the results are interpreted:
   - Deming regression
   - Nonparametric regression
   - Ordinary least-squares regression
   - Regression
9. Describe the calibration hierarchy, including the tracing of values of routine clinical chemistry measurements to a primary reference, how the values are obtained, and the methods involved; draw a calibration hierarchy given a specific analyte.
10. Discuss the concept of uncertainty in relation to clinical laboratory results, including the components of the standard uncertainty formula and two ways in which uncertainty is assessed.
11. Given appropriate values, state the formula and calculate the following:
    - Coefficient of variation
    - Coefficient of variation percent
    - Deming regression
    - Linear regression
    - Population mean
    - Precision analyses
    - Standard deviation
    - Standard uncertainty

- Likelihood ratio
- Odds ratio
- Predictive value
- Prevalence
- Receiver operating characteristic curve
- Sensitivity
- Specificity
- True/false positive
- True/false negative

12. State the formulas for and calculate, given appropriate information, the following: sensitivity, specificity, predictive value for positive/negative tests (posterior probabilities), odds ratio, and positive/negative likelihood ratio.
13. State the relationship between high sensitivity and false negatives; state the relationship between high specificity and false positives.
14. Compare dichotomous and continuous tests; include definition, sensitivity/specificity, and a clinical example of each type of test.
15. State how the predictive value of a laboratory test (posterior probability) is affected by prevalence.
16. Construct and interpret a receiver operating characteristic curve using data from a diagnostic test study.
17. Describe the added value of combination testing as it is used in the clinical laboratory; include examples, diagnostic usefulness, and associated problems.

Accuracy of analytical method Closeness of agreement of a single measurement with "true value."

Analyte The substance being analyzed in an analytical procedure.

Analytical measurement range The analyte concentration range over which measurements are within the declared tolerances for imprecision and bias; also referred to as *reportable range.*

Analytical sensitivity The ability of an analytical method to assess small variations in the concentration of analyte.

Analytical specificity The ability of an assay procedure to determine specifically the concentration of the target analyte in the presence of potentially interfering substances or factors in the sample matrix.

Assay comparison Comparison of measurements by two methods that is carried out objectively using statistical procedures and graphics displays.

Bias In an analytical method, the difference between the average value and the true value that is expressed numerically and is inversely related to the trueness.

Calibration In relation to analytical methods, a function that describes the relationship between instrument signal and concentration of analyte.

Coefficient of variation Relative standard deviation.

Commutability The equivalence of the mathematical relationships between the results of different measurement procedures for a reference material and for representative samples from healthy and diseased individuals.

Correlation coefficient Measure of association between two variables.

Deming regression Least-squares regression analysis taking measurement errors in both variables into account.

Difference plot A bias plot that shows the dispersion of observed differences between the measurements of two methods as a function of the average concentration of the measurements; also referred to as a *Bland-Altman plot.*

Error model A model of the error structure.

Frequency distribution A distribution of the frequency (ordinate) as a function of the variable value (abscissa), that is, a histogram of absolute or relative frequencies.

Gaussian probability distribution Bell shaped relative frequency distribution described under basic statistics.

Likelihood ratio The probability of occurrence of a specific test value given that the disease is present divided by the probability of the same test value if the disease was absent.

Linearity Range of values for which there is a linear relationship between concentration and signal.

Limit of detection An assay characteristic defined as the lowest value that significantly exceeds the measurements of a blank sample.

Matrix In relation to analytical methods, human serum that contains analytes.

Mean Arithmetic average of variables. See Basic Statistics.

Median Equal to the 50th percentile of a set of variables. See Basic Statistics.

Negative predictive value The proportion of subjects with a negative test who do not have the disease.

Odds ratio The probability of the presence of a specific disease divided by the probability of its absence.

Ordinary least-squares regression (OLR) analysis A method used to estimate the unknown parameters in a linear regression assessment performed to minimize the sum of squared vertical distances between observed responses and responses predicted by linear approximation.

Population In relation to analytical methods, the complete set of all observations that might occur as the result of performing a particular procedure according to specified conditions.

Positive predictive value The proportion of subjects with a positive test who have the disease.

Precision The closeness of agreement between independent results of measurements obtained under stipulated conditions. Usually expressed as the standard deviation.

Prevalence The frequency of disease in the population examined.

Random error Error that arises from imprecision of measurement of the type that is described by a Gaussian distribution (e.g., caused by pipetting variability, signal variability).

Random sample A random sample from a population is one in which each member has an equal chance of being selected.

Receiver operating characteristic curve A graph of sensitivity versus 1—specificity for all possible cutoff values of a diagnostic test; used to estimate sensitivity and specificity for various decision cutoffs.

Reference measurement procedure A procedure of highest analytical quality that has been shown to yield values having an uncertainty of measurement commensurate with its intended use, especially in assessing the trueness of other measurement procedures for the same quantity and in characterizing reference materials.

Regression analysis A statistical analysis that compares measurement relations between two analytical methods.

Repeatability Closeness of agreement between results of successive measurements carried out under the same conditions (i.e., corresponding to within-run precision).

Reproducibility Closeness of agreement between results of successive measurements carried out under changed conditions (e.g., corresponding to between-runs).

Sample A finite set of variables drawn from an infinite population of variables.

Sensitivity The proportion of subjects with disease who have a positive laboratory test result.

Specificity The proportion of subjects without disease who have a negative laboratory test result.

Standard deviation Square root of sum of squared deviations from the mean divided by number of variables minus one. See under Basic Statistics.

Student *t* distribution Distribution related to the Gaussian distribution given a limited sample size. See under Basic Statistics.

Systematic error Error in measurement that arises from calibration bias or nonspecificity of an assay and, in the course of a number of analyses of the same analyte, remains constant (y-intercept deviation from zero) or varies in a proportional way (slope deviation from unity) based on the analyte concentration.

Traceability In relation to analytical methods, a concept based on a chain of comparisons of measurements that lead to a known reference value done to ensure reasonable agreement between measurements of routine methods.

Trueness A qualitative term that describes the closeness of agreement between the average value obtained from a large series of results of measurements and a true value.

Uncertainty A parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measure; more briefly, uncertainty is a parameter characterizing the range of values within which the value of the quantity being measured is expected to lie.

ASSAY SELECTION OVERVIEW

The introduction of new or revised laboratory tests, markers, or assays is a common occurrence in the clinical laboratory. Test selection and evaluation are key steps in the process of implementing new measurement procedures (Fig. 2.1).

Evaluation of tests, markers, or assays in the clinical laboratory is influenced strongly by guidelines and accreditation or other regulatory standards. The Clinical and Laboratory Standards Institute (CLSI) has published a series of consensus protocols for clinical chemistry laboratories and manufacturers to follow when evaluating methods (http://www.clsi.org). The International Organization for Standardization (ISO) has also developed several documents related to method evaluation (ISOs). In addition, meeting laboratory accreditation requirements has become an important aspect in most countries. Abbreviations are listed in Box 2.1.

Analytical Performance Criteria

In evaluation of a laboratory test, (1) trueness (formerly termed accuracy), (2) precision, (3) analytical range, (4) detection limit, and (5) analytical specificity are of prime importance. The sections in this chapter on laboratory test evaluation and comparison contain detailed outlines of these concepts. Estimated test performance parameters should be related to analytical performance specifications that ensure acceptable clinical use of the test and its results. For more details related to the recommended models for setting analytical performance specifications, readers are referred to Chapters 4 and 7. From a practical point of view, the "ruggedness" of the test in routine use is of importance, and reliable performance, when used by different operators and with different batches of reagents over long time periods, is essential.

BASIC STATISTICS

In this section, fundamental statistical concepts and techniques are introduced in the context of typical analytical investigations. The basic concepts of (1) populations, (2) samples, (3) parameters, (4) statistics, and (5) probability distributions are defined and illustrated. Two important probability distributions—Gaussian and Student *t*—are introduced and discussed.

Frequency Distribution

A graphical device for displaying a large set of laboratory test results is the **frequency distribution**, also called a *histogram*.

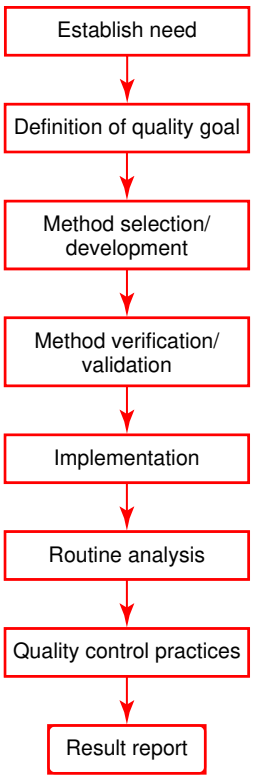


Fig. 2.1 A flow diagram that illustrates the process of introducing a new assay into routine use.

BOX 2.1 Abbreviations

CI	Confidence interval
CV	Coefficient of variation (= SD/x, where x is the concentration)
CV%	= CV × 100%
CV_A	Analytical coefficient of variation
CV_{RB}	Sample-related random bias coefficient of variation
ISO	International Organization for Standardization
OLR	Ordinary least-squares regression analysis
SD	Standard deviation
SEM	Standard error of the mean $\left(=SD/\sqrt{N}\right)$
SD_A	Analytical standard deviation
SD_{RB}	Sample-related random bias standard deviation
x_m	Mean
x_{mw}	Weighted mean

Fig. 2.2 shows a frequency distribution displaying the results of serum gamma-glutamyltransferase (GGT) measurements of 100 apparently healthy 20- to 29-year-old men. The frequency distribution is constructed by dividing the measurement scale into cells of equal width; counting the number, n_i, of values that fall within each cell; and drawing a rectangle above each cell whose area (and height because the cell widths are all equal) is proportional to n_i. In this example, the selected cells were 5 to 9, 10 to 14, 15 to 19, 20 to 24, 25 to 29, and so on, with 60 to 64 being the last cell (range of values, 5 to 64 U/L). The ordinate axis of the frequency distribution

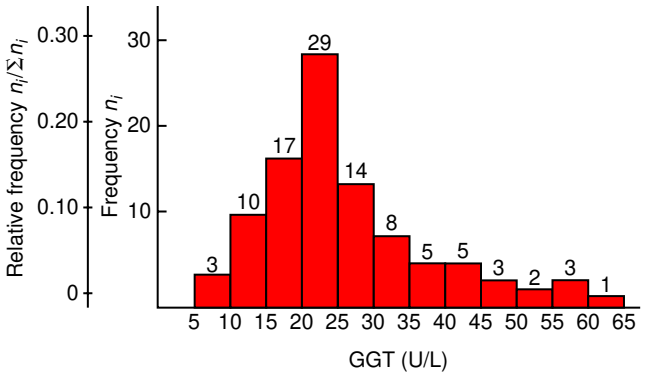


Fig. 2.2 Frequency distribution of 100 gamma-glutamyltransferase *(GGT)* values.

gives the number of values falling within each cell. When this number is divided by the total number of values in the data set, the relative frequency in each cell is obtained.

Often, the position of the value for an individual within a distribution of values is useful medically. The *nonparametric* approach can be used to determine directly the *percentile* of a given subject. Having ranked N subjects according to their values, the n-percentile, $Perc_n$, may be estimated as the value of the $[N(n/100) + 0.5]$ ordered observation. In the case of a non-integer value, interpolation is carried out between neighbor values. The 50th percentile is the median of the distribution.

Population and Sample

It is useful to obtain information and draw conclusions about the characteristics of the test results for one or more target populations. In the GGT example, interest is focused on the location and spread of the population of GGT values for 20- to 29-year-old healthy men. Thus a working definition of a **population** is the complete set of all observations that might occur as a result of performing a particular procedure according to specified conditions.

Most target populations of interest in clinical chemistry are in principle very large (millions of individuals) and so are impossible to study in their entirety. Usually a subgroup of observations is taken from the population as a basis for forming conclusions about population characteristics. The group of observations that has been selected from the population is called a **sample**. For example, the 100 GGT values make up a sample from a respective target population. However, a sample is used to study the characteristics of a population only if it has been properly selected. For instance, if the analyst is interested in the population of GGT values over various lots of materials and some time period, the sample must be selected to be representative of these factors, as well as of age, sex, and health factors of the individuals in the targeted population. Consequently, exact specification of the target population(s) is necessary before a plan for obtaining the sample(s) can be designed. In this chapter, a sample is also used as a specimen, depending on the context.

Probability and Probability Distributions

Consider again the frequency distribution in Fig. 2.2. In addition to the general location and spread of the GGT

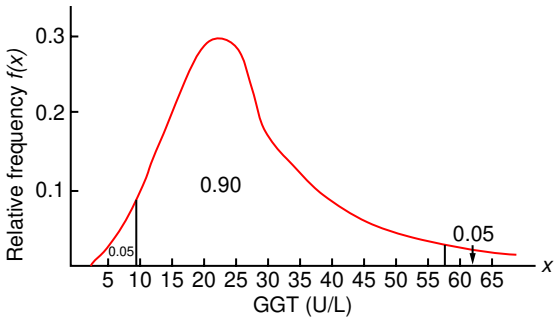


Fig. 2.3 Population frequency distribution of gamma-glutamyltransferase *(GGT)* values.

determinations, other useful information can be easily extracted from this frequency distribution. For instance, 96% (96 of 100) of the determinations are less than 55 U/L, and 91% (91 of 100) are greater than or equal to 10 but less than 50 U/L. Because the cell interval is 5 U/L in this example, statements such as these can be made only to the nearest 5 U/L. A larger sample would allow a smaller cell interval and more refined statements. For a sufficiently large sample, the cell interval can be made so small that the frequency distribution can be approximated by a continuous, smooth curve, similar to that shown in Fig. 2.3. In fact, if the sample is large enough, this can be considered a close representation of the "true" target *population frequency distribution*. In general, the functional form of the population frequency distribution curve of a variable x is denoted by $f(x)$.

The population frequency distribution allows us to make probability statements about the GGT of a randomly selected member of the population of healthy 20- to 29-year-old men. For example, the probability $\Pr(x > x_a)$ that the GGT value x of a randomly selected 20- to 29-year-old healthy man is greater than some particular value x_a is equal to the area under the population frequency distribution to the right of x_a. If $x_a = 58$, then from Fig. 2.3, $\Pr(x > 58) = 0.05$. Similarly, the probability $\Pr(x_a < x < x_b)$ that x is greater than x_a but less than x_b is equal to the area under the population frequency distribution between x_a and x_b. For example, if $x_a = 9$ and $x_b = 58$, then from Fig. 2.3, $\Pr(9 < x < 58) = 0.90$. Because the population frequency distribution provides all information related to probabilities of a randomly selected member of the population, it is called the probability distribution of the population. Although the true probability distribution is never exactly known in practice, it can be approximated with a large sample of observations, that is, test results.

Parameters: Descriptive Measures of a Population

Any population of values can be described by measures of its characteristics. A *parameter* is a constant that describes some particular characteristic of a population. Although most populations of interest in analytical work are infinite in size, for the following definitions, the population will be considered to be of finite size N, where N is very large.

One important characteristic of a population is its *central location*. The parameter most commonly used to describe the central location of a population of N values is the *population mean* (μ):

$$\mu = \frac{\sum x_i}{N}$$

An alternative parameter that indicates the central tendency of a population is the **median**, which is defined as the 50th percentile, Perc_{50}.

Another important characteristic is the *dispersion* of values about the population mean. A parameter very useful in describing this dispersion of a population of N values is the *population variance* σ^2 (sigma squared):

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

The *population standard deviation (SD)* σ, the positive square root of the population variance, is a parameter frequently used to describe the population dispersion in the same units (e.g., mg/dL) as the population values. For a Gaussian distribution, 95% of the population of values are located within the mean $\pm 1.96\ \sigma$. If a distribution is non-Gaussian (e.g., asymmetric), an alternative measure of dispersion based on the percentiles may be more appropriate, such as the distance between the 25th and 75th percentiles (the interquartile interval).

Statistics: Descriptive Measures of the Sample

As noted earlier, clinical chemists usually have at hand only a sample of observations (i.e., test results) from the overarching targeted population. A *statistic* is a value calculated from the observations in a sample to estimate a particular characteristic of the target population. As introduced earlier, the sample mean x_m is the arithmetical average of a sample, which is an estimate of μ. Likewise, the sample **standard deviation** (SD) is an estimate of σ, and the **coefficient of variation** (CV) is the ratio of the SD to the mean multiplied by 100%. The equations used to calculate x_m, SD, and CV, respectively, are as follows:

$$x_m = \frac{\sum x_i}{N}$$

$$\text{SD} = \sqrt{\frac{\sum (x_i - x_m)^2}{N-1}} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N-1}}$$

$$\text{CV} = \frac{\text{SD}}{x_m} \times 100\%$$

where x_i is an individual measurement and N is the number of sample measurements.

The SD is an estimate of the dispersion of the distribution. In addition, from the SD, an estimate of the uncertainty of x_m can be derived as an estimate of μ (see later discussion).

Random Sampling

A random sample of individuals from a target population is one in which each member of the population has an equal chance of being selected. A **random sample** is one in which each member of the sample can be considered to be a random

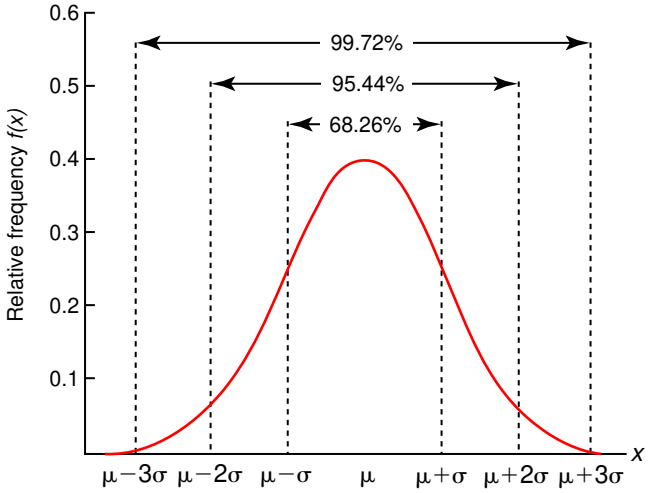


Fig. 2.4 The Gaussian probability distribution.

selection from the target population. Although much of statistical analysis and interpretation depends on the assumption of a random sample from some population, actual data collection often does not satisfy this assumption. In particular, for sequentially generated data, it is often true that observations adjacent to each other tend to be more alike than observations separated in time.

Gaussian Probability Distribution

The **Gaussian probability distribution**, illustrated in Fig. 2.4, is of fundamental importance in statistics for several reasons. As mentioned earlier, a particular test result x will not usually be equal to the true value μ of the specimen being measured. Rather, associated with this particular test result x will be a particular measurement error $\varepsilon = x - \mu$, which is the result of many contributing sources of error. Pure measurement errors tend to follow a probability distribution similar to that shown in Fig. 2.4, where the errors are symmetrically distributed, with smaller errors occurring more frequently than larger ones, and with an expected value of 0. This important fact is known as the central limit effect for distribution of errors: if a measurement error ε is the sum of many independent sources of error, such as $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$, several of which are major contributors, the probability distribution of the measurement error ε will tend to be Gaussian as the number of sources of error becomes large.

Another reason for the importance of the Gaussian probability distribution is that many statistical procedures are based on the assumption of a Gaussian distribution of values; this approach is commonly referred to as *parametric*. Furthermore, these procedures usually are not seriously invalidated by departures from this assumption. Finally, the magnitude of the uncertainty associated with sample statistics can be ascertained based on the fact that many sample statistics computed from large samples have a Gaussian probability distribution.

The Gaussian probability distribution is completely characterized by its mean μ and its variance σ^2. The notation $N(\mu, \sigma^2)$ is often used for the distribution of a variable that is Gaussian with mean μ and variance σ^2. Probability statements

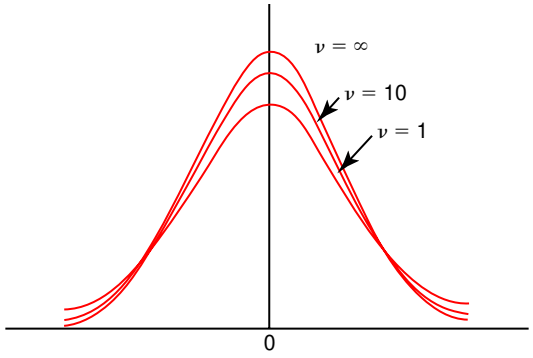


Fig. 2.5 The t distribution for $\nu = 1$, 10, and ∞.

about a variable x that follows an $N(\mu, \sigma^2)$ distribution are usually made by considering the variable z,

$$z = \frac{x - \mu}{\sigma}$$

which is called the *standard Gaussian variable*. The variable z has a Gaussian probability distribution with $\mu = 0$ and $\sigma^2 = 1$, that is, z is $N(0, 1)$. The probably that x is within 2σ of μ [i.e., $\Pr(|x - \mu| < 2\sigma) =$] is 0.9544. Most computer spreadsheet programs can calculate probabilities for all values of z.

Student *t* Probability Distribution

To determine probabilities associated with a Gaussian distribution, it is necessary to know the population SD σ. In actual practice, σ is often unknown, so z cannot be calculated. However, if a random sample can be taken from the Gaussian population, the sample SD can be calculated, by substituting SD for σ, and computing the value t:

$$t = \frac{x - \mu}{\mathrm{SD}}$$

Under these conditions, the variable t has a probability distribution called the **Student *t* distribution**. The t distribution is a family of distributions depending on the degrees of freedom ν $(= N - 1)$ for the sample SD. Several t distributions from this family are shown in Fig. 2.5. When the size of the sample and the degrees of freedom for SD are infinite, there is no uncertainty in SD, so the t distribution is identical to the standard Gaussian distribution. However, when the sample size is small, the uncertainty in SD causes the t distribution to have greater dispersion and heavier tails than the standard Gaussian distribution, as illustrated in Fig. 2.5. At sample sizes above 30, the difference between the t distribution and the Gaussian distribution becomes relatively small and can usually be neglected. Most computer spreadsheet programs can calculate probabilities for all values of t, given the degrees of freedom for SD.

The Student t distribution is commonly used in significance tests, such as comparison of sample means, or in testing conducted if a regression slope differs significantly from 1. Descriptions of these tests can be found in statistics textbooks. Another important application is the estimation of confidence intervals (CIs). CIs are intervals that indicate the uncertainty of a given sample estimate. For example, it can be proved that $X_m \pm t_{alpha}\ (\mathrm{SD}/N^{0.5})$ provides an approximate

2*alpha*-CI for the mean. A common value for *alpha* is 0.025 or 2.5%, which thus results in a 0.95% or 95% CI. Given sample sizes of 30 or higher, t_{alpha} is about 2. ($SD/N^{0.5}$) is called the standard error (SE) of the mean. A CI should be interpreted as follows. Suppose a sampling experiment of drawing 30 observations from a Gaussian population of values is repeated 100 times, and in each case, the 95% CI of the mean is calculated as described. Then, in 95% of the drawings, the true mean μ is included in the 95% CI. According to the central limit theorem, distributions of mean values converge toward the Gaussian distribution irrespective of the primary type of distribution of *x*. This means that the 95% CI is a robust estimate only minimally influenced by deviations from the Gaussian distribution. In the same way, the *t*-test is robust toward deviations from normality.

Nonparametric Statistics

Distribution-free statistics, often called nonparametric statistics, provides an alternative to parametric statistical procedures that assume data to have Gaussian distributions. For example, distributions of reference values are often skewed and so do not conform to the Gaussian distribution (see Chapter 5 on reference intervals). Formally, to judge whether a distribution is Gaussian or not, a goodness of fit test should be conducted. A commonly used test is the Kolmogorov-Smirnov test, in which the shape of the sample distribution is compared with the shape presumed for a Gaussian distribution. If the difference exceeds a given critical value, the hypothesis of a Gaussian distribution is rejected, and it is then appropriate to apply nonparametric statistics. A special problem is the occurrence of outliers, (i.e., single measurements highly deviating from the remaining measurements). Outliers may rely on biological factors and so be of real significance (e.g., in the context of estimating reference intervals) or they may be related to clerical errors. Special tests exist for handling outliers.

Given that a distribution is non-Gaussian, it is appropriate to apply nonparametric descriptive statistics based on the percentile or quantile concept. The *n*-percentile, $Perc_n$, of a sample of *N* values may be estimated as the value of the [$N(n/100)$ + 0.5] ordered observation. In the case of a noninteger value, interpolation is carried out between neighbor values. The median is the 50th percentile, which is used as a measure of the center of the distribution. For the GGT example, we would order the $N = 100$ values according to size. The median or 50th percentile is then the value of the [100(50/100) + 0.5 = 50.5] ordered observation (the interpolated value between the 50th and 51st ordered values). The 2.5th and 97.5th percentiles are values of the [100(2.5/100) + 0.5 = 3] and [100(97.5/100) + 0.5 = 98] ordered observations, respectively. When a 95% reference interval is estimated, a nonparametric approach is often preferable because many distributions of reference values are asymmetric. Generally, distributions based on the many biological sources of variation are often non-Gaussian compared with distributions of pure measurement errors that usually are Gaussian.

The nonparametric counterpart to the *t*-test is the Mann-Whitney test, which provides a significance test for the difference between median values of the two groups to be compared, given the same shape of the distributions. When there are more than two groups, the Kruskal-Wallis test can be applied.

Categorical Variables

When dealing with qualitative tests and in the context of evaluating diagnostic testing, categorical variables that only take the value positive or negative come into play. The performance is here given as proportions or percentages, which are proportions multiplied by 100. For example, the diagnostic **sensitivity** of a test is the proportion of diseased subjects who have a positive result. Having tested, for example, 100 patients, 80 might have had a positive test result. The sensitivity then is 0.8% or 80%. Exact estimates of the uncertainty can be derived from the so-called binomial distribution, but for practical purposes, an approximate expression for the 95% CI is usually applied as the estimated proportion $P \pm 2SE$, where the SE in this context is derived as:

$$SE = [P(1-P)/N]^{0.5}$$

where *P* is here a proportion and not a percentage. In the example, the SE equals 0.0016 and so the 95% CI is 0.77 to 0.83 or 77% to 83%. The applied approximate formula for the SE is regarded as reasonably valid when *NP* and $N(1-P)$ both are equal to or higher than 5.

TECHNICAL VALIDITY OF ANALYTICAL ASSAYS

This section defines the basic concepts used in this chapter: (1) **calibration**, (2) trueness and accuracy, (3) precision, (4) linearity, (5) **limit of detection** (LOD), (6) limit of quantification, (7) **specificity**, and (8) others.

Calibration

The calibration function is the relation between instrument signal *(y)* and concentration of **analyte** *(x)*, that is,

$$y = f(x)$$

The inverse of this function, also called the measuring function, yields the concentration from response:

$$x = f^{-1}(y)$$

This relationship is established by measurement of samples with known quantities of analyte (calibrators). Solutions of pure chemical standards should be distinguished from samples with known quantities of analyte present in the typical **matrix** that is to be measured (e.g., human serum). The first situation applies typically to a reference measurement procedure that is not influenced by matrix effects; the second case corresponds typically to a routine method that often is influenced by matrix components and so preferably is calibrated using the relevant matrix. Calibration functions may be linear or curved and, in

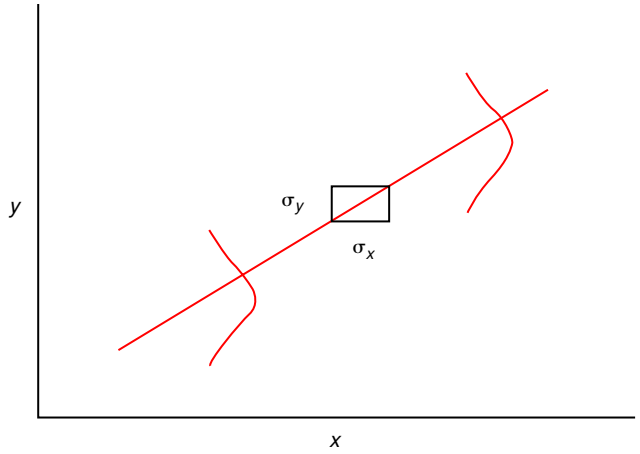


Fig. 2.6 Relation between concentration *(x)* and signal response *(y)* for a linear calibration function. The dispersion in signal response (σ_y) is projected onto the *x*-axis and is called assay imprecision [σ_x (=σ_A)].

TABLE 2.1 Overview of Qualitative Terms and Quantitative Measures Related to Method Performance

Qualitative Concept	Quantitative Measure
Trueness	*Bias*
Closeness of agreement of mean value with "true value"	A measure of the systematic error
Precision	*Imprecision (SD)*
Repeatability (within run)	A measure of the dispersion of random errors
Intermediate precision (long term)	
Reproducibility (interlaboratory)	
Accuracy	*Error of measurement*
Closeness of agreement of a single measurement with "true value"	Comprises both random and systematic influences

SD, Standard deviation.

the case of immunoassays, may often take a special form (e.g., modeled by the four-parameter logistic curve). An alternative, model-free approach is to estimate a smoothed spline curve, which often is performed for immunoassays. If the assumed calibration function does not correctly reflect the true relationship between instrument response and analyte concentration, a **systematic error** or **bias** is likely to be associated with the analytical method.

The precision of the analytical method depends on the stability of the instrument response for a given quantity of analyte. In principle, a random dispersion of instrument signal (vertical direction) at a given true concentration transforms into dispersion on the measurement scale (horizontal direction), as is shown schematically (Fig. 2.6). If the calibration function is linear and the imprecision of the signal response is the same over the analytical measurement range, the analytical SD (SD_A) of the method tends to be constant over the analytical measurement range (see Fig. 2.6). If the imprecision increases proportionally to the signal response, the analytical SD of the method tends to increase proportionally to the concentration *(x)*, which means that the *relative* imprecision (CV = SD/*x*) may be constant over the analytical measurement range assuming that the intercept of the calibration line is zero.

Trueness and Accuracy

Trueness of measurements is defined as closeness of agreement between the average value obtained from a large series of results of measurements and the true value.

The difference between the average value (strictly, the mathematical expectation) and the true value is the *bias*, which is expressed numerically and so is inversely related to the trueness. *Trueness* in itself is a qualitative term that can be expressed, for example, as low, medium, or high. From a theoretical point of view, the exact true value for a clinical sample is not available; instead, an "accepted reference value" is used, which is the "true" value that can be determined in practice. Trueness can be evaluated by comparison of measurements by the new test and by some preselected reference measurement procedure, both on the same sample or individuals.

The ISO has introduced the trueness expression as a replacement for the term *accuracy*, which now has gained a slightly different meaning. **Accuracy** is the closeness of agreement between the result of a measurement and a true concentration of the analyte. Accuracy thus is influenced by both bias and imprecision and in this way reflects the total error. Accuracy, which in itself is a qualitative term, is inversely related to the "uncertainty" of measurement, which can be quantified as described later (Table 2.1).

In relation to trueness, the concepts *recovery*, *drift*, and *carryover* may also be considered. *Recovery* is the fraction or percentage increase in concentration that is measured in relation to the amount added. Recovery experiments are typically carried out in the field of drug analysis. It is useful to distinguish between *extraction recovery*, which often is interpreted as the fraction of compound that is carried through an extraction process, and the recovery measured by the entire analytical procedure, in which the addition of an internal standard compensates for losses in the extraction procedure. A recovery close to 100% is a prerequisite for a high degree of trueness, but it does not ensure unbiased results because possible nonspecificity against matrix components (e.g., an interfering substance) is not detected in a recovery experiment. *Drift* is caused by instrument or reagent instability over time, so that calibration becomes gradually biased. Assay *carryover* also must be close to zero to ensure unbiased results.

Precision

Precision has been defined as the closeness of agreement between independent results of measurements obtained under stipulated conditions. The degree of precision is usually derived from statistical measures of imprecision, such as SD or CV (CV = SD/*x*, where *x* is the measurement concentration), which is inversely related to precision. Imprecision of measurements is solely related to the **random error** of measurements and has no relation to the trueness of measurements.

Precision is specified as follows:

Repeatability: **Repeatability** is the closeness of agreement between results of successive measurements carried out under the same conditions (i.e., corresponding to within-run precision).

Reproducibility: **Reproducibility** is the closeness of agreement between results of measurements performed under changed conditions of measurements (e.g., time, operators, calibrators, reagent lots). Two specifications of reproducibility are often used: total or between-run precision in the laboratory, often termed *intermediate precision*, and inter-laboratory precision (e.g., as observed in external quality assessment schemes [EQAS]) (see Table 2.1).

The total SD (σ_T) may be divided into within-run and between-run components using the principle of analysis of variance of components (variance is the squared SD):

$$\sigma_T^2 = \sigma_{\text{Within-run}}^2 + \sigma_{\text{Between-run}}^2$$

It is not always clear in clinical chemistry publications what is meant by "between-run" variation. Some authors use the term to refer to the total variation of an assay, but others apply the term *between-run variance component* as defined earlier.

In laboratory studies of analytical variation, estimates of imprecision are obtained. It is important to have an adequate number so that analytical variation is not underestimated. Commonly, the number 20 is given as a reasonable number of observations (e.g., suggested in the CLSI guideline for manufacturers). To verify method precision by users, it has been recommended to run internal QC samples for 5 consecutive days in 5 replicates.

To estimate both the within-run imprecision and the total imprecision, a common approach is to measure duplicate control samples in a series of runs. Suppose, for example, that a control is measured in duplicate for 20 runs, in which case 20 observations are present with respect to both components. The dispersion of the means (x_m) of the duplicates is given as follows:

$$\sigma_{x_m}^2 = \frac{\sigma_{\text{Within-run}}^2}{2} + \sigma_{\text{Between-run}}^2$$

From the 20 sets of duplicates, the within-run SD can be derived using the following formula:

$$\text{SD}_{\text{Within-run}} = \left[\Sigma \frac{d_i^2}{(2 \times 20)}\right]^{0.5}$$

where d_i refers to the difference between the ith set of duplicates. When SDs are estimated, the concept degrees of freedom *(df)* is used. In a simple situation, the number of degrees of freedom equals $N - 1$. For N duplicates, the number of degrees of freedom is $N(2 - 1) = N$. Thus both variance components are derived in this way. The advantage of this approach is that the within-run estimate is based on several runs, so that an average estimate is obtained rather than only an estimate for one particular run if all 20 observations had been obtained in the same run.

Generally, the estimate of the imprecision improves as more observations become available. Exact confidence limits for the SD can be derived from the χ^2 distribution. A graphical display of 95% CIs at various sample sizes is shown in Fig. 2.7. For example, suppose we have estimated the imprecision as an SD of 5.0 on the basis of $N = 20$ observations. From the figure, the 95% CI extends from about 0.75 × 5.0 to about 1.45 × 5.0, i.e., from 3.8 to 7.3.

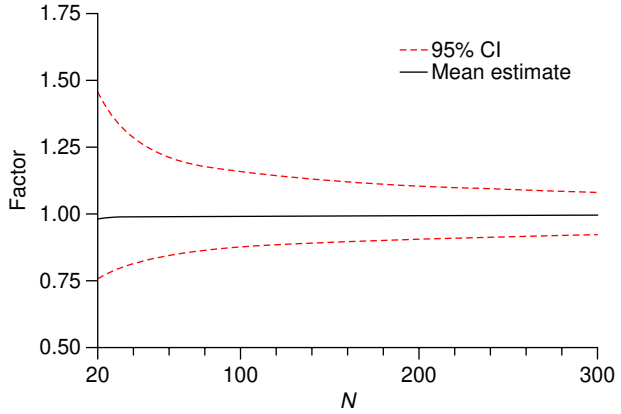


Fig. 2.7 Relation between factors indicating the 95% confidence intervals *(CIs)* of standard deviations (SDs) and the sample size. The true SD is 1, and the solid line indicates the mean estimate, which is slightly downward biased at small sample sizes.

Precision often depends on the concentration of analyte being considered. A presentation of precision as a function of analyte concentration is the precision profile, which usually is plotted in terms of the SD or the CV as a function of analyte concentration (Fig. 2.8).

Linearity

Linearity refers to the relationship between measured and expected values over the analytical measurement range. Linearity may be considered in relation to actual or relative analyte concentrations. In the latter case, a dilution series of a sample may be examined. This dilution series examines whether the measured concentration changes as expected according to the proportional relationship between samples introduced by the dilution factor. Dilution is usually carried out with an appropriate sample matrix (e.g., human serum [individual or pooled serum] or a verified sample diluent).

Evaluation of linearity can be performed visually or by statistical tests. When repeated measurements are available at each concentration, the random variation between measurements and the variation around an estimated regression line may be evaluated by an F-test. When significant nonlinearity is found, it may be useful to explore nonlinear alternatives to the linear regression line (i.e., polynomials of higher degrees). Another approach is to assess the residuals of an estimated regression line. An additional consideration for evaluating proportional concentration relationships is whether an estimated regression line passes through zero or not.

Analytical Measurement Range and Limits of Quantification

The **analytical measurement range** (measuring interval, reportable range) is the analyte concentration range over which measurements are within the declared tolerances for imprecision and bias of the method. Taking drug assays as an example, there exist (arbitrary) requirements of a CV% of less than 15% and a bias of less than 15%. The measurement range then extends from the lowest concentration (lower limit of quantification [LLOQ]) to the highest concentration (upper limit of

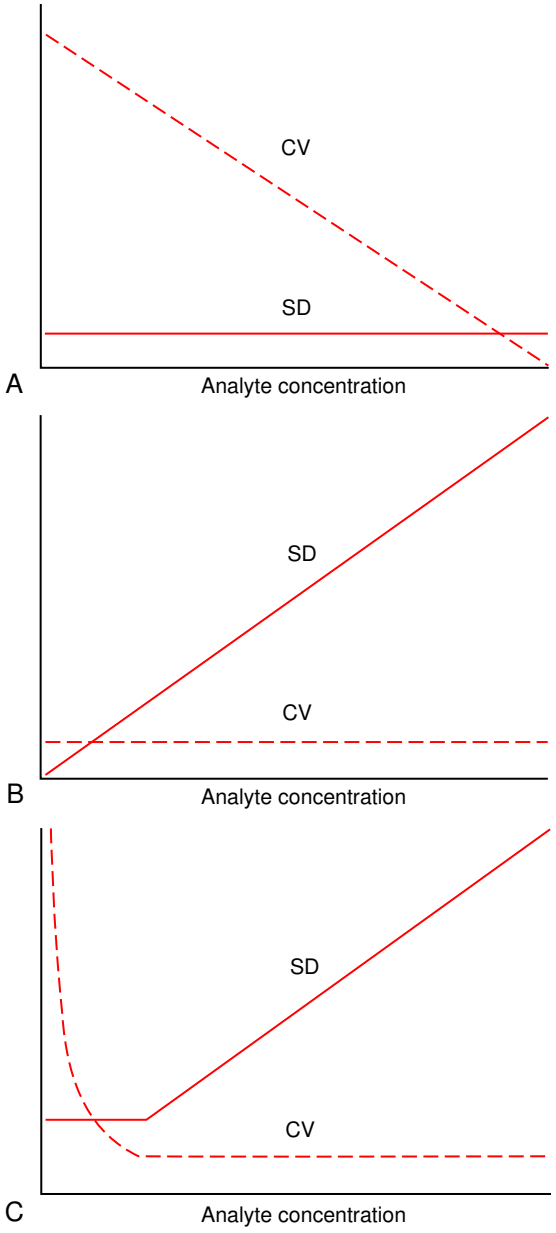


Fig. 2.8 Relations between analyte concentration and standard deviation *(SD)*/coefficient of variation *(CV)*. (A) The SD is constant, so that the CV varies inversely with the analyte concentration. (B) The CV is constant because of a proportional relationship between concentration and SD. (C) A mixed situation with constant SD in the low range and a proportional relationship in the rest of the analytical measurement range.

quantification [ULOQ]) for which these performance specifications are fulfilled. The LLOQ is medically important for many analytes, e.g., thyroid-stimulating hormone (TSH), where low TSH results are useful for the diagnosis of hyperthyroidism.

The LOD may be defined as the lowest value that significantly exceeds the measurements of a blank sample. Thus the limit has been estimated on the basis of repeated measurements of a blank sample and has been *reported* as the mean plus 2 or 3 SDs of the blank measurements. In the interval from LOD up to LLOQ, a result should be reported as "detected" but not provided as a quantitative result.

The LLOQ of an assay should not be confused with analytical sensitivity. The **analytical sensitivity** is defined as ability of an analytical method to assess small differences in the concentration of analyte. The analytical sensitivity depends on the precision of the method. The smallest difference that will be statistically significant equals $2 \times \sqrt{2 \times SD_A}$ at a 5% significance level.

Analytical Specificity and Interference

Analytical specificity is the ability of an assay procedure to determine the concentration of the target analyte without influence from potentially interfering substances or factors in the sample matrix (e.g., hyperlipemia, hemolysis, bilirubin, antibodies, other metabolic molecules, degradation products of the analyte, exogenous substances, anticoagulants).

POINTS TO REMEMBER

- Technical validation of analytical methods focuses on (1) calibration, (2) trueness and accuracy, (3) precision, (4) linearity, (5) LOD, (6) limit of quantification, (7) specificity, and (8) others.
- The difference between the average measured value and the true value is the bias, which can be evaluated by comparison of measurements by the new test and by some preselected reference measurement procedure, both on the same sample or individuals.
- The degree of precision is usually expressed on the basis of statistical measures of imprecision, such as SD or CV (CV = SD/x, where x is the measurement concentration).
- The measurement range extends from the lowest concentration (LLOQ) to the highest concentration (ULOQ) for which the analytical performance specifications (imprecision, bias) are fulfilled.
- Analytical specificity is the ability of an assay procedure to determine the concentration of the target analyte without influence from potentially interfering substances or factors in the sample matrix.

ASSAY COMPARISON

Comparison of measurements by two assays can be carried out by parallel measurements of a set of patient samples. To prevent artificial matrix-induced differences, fresh patient samples are the optimal material. A nearly even distribution of values over the analytical measurement range is also preferable. In an ordinary laboratory, comparison of two routine assays is the most frequently occurring situation. Less commonly, comparison of a routine assay with a reference measurement procedure is undertaken. When two routine assays are compared, it is not possible to establish that one set of measurements is the correct one. Rather, the question is whether the new assay can replace the existing one without a systematic change in result values. To address this question, a statistical procedure with graphics display should be applied. A difference (bias) plot, which shows differences as a function of the average concentration of measurements (Bland-Altman plot), or a **regression analysis**.

Basic Error Model

The occurrence of measurement errors is related to the performance characteristics of the assay. It is important to distinguish between pure, random measurement errors, which are present in all measurement procedures, and errors related to incorrect calibration and nonspecificity of the assay. Whereas a reference measurement procedure is associated only with pure random error, a routine method, additionally, is likely to have some bias related to errors in calibration and limitations with regard to specificity. Whereas an erroneous calibration function gives rise to a systematic error, nonspecificity gives an error that typically varies from sample to sample. The error related to nonspecificity thus has a random character, but in contrast to the pure measurement error, it cannot be reduced by repeated measurements of a sample. Although errors related to nonspecificity for a group of samples look like random errors, for the individual sample, this type of error is a bias. Because this bias varies from sample to sample, it has been called a *sample-related random bias.* In the following section, the various error components are incorporated into a formal **error model**.

Measured Value, Target Value, Modified Target Value, and True Value

Taking into account that an analytical method measures analyte concentrations with some random measurement error, it is necessary to distinguish between the actual, measured value and the average result obtained if the given sample was measured an infinite number of times. If the assay is a reference assay without bias and nonspecificity, the following simple relationship holds:

$$x_i = X_{\text{True}i} + \varepsilon_i$$

where x_i represents the measured value, $X_{\text{True}i}$ is the average value for an infinite number of measurements, and ε_i is the deviation of the measured value from the average value. If the sample was measured repeatedly, the average of ε_i would be zero and the SD would equal the analytical SD (σ_A) of the reference measurement procedure. Pure, random measurement error will usually be Gaussian distributed.

In the case of a routine assay, the relationship between the measured value for a sample and the true value becomes more complicated:

$$x_i = X_{\text{True}i} + \text{Cal-Bias} + \text{Random-Bias}_i + \varepsilon_i$$

The *Cal-Bias* term (calibration bias) is a systematic error related to the calibration of the method. This systematic error may be a constant for all measurements corresponding to an offset error, or it may be a function of the analyte concentration (e.g., corresponding to a slope deviation in the case of a linear calibration function). The *Random-Bias*$_i$ term is a bias that is specific for a given sample related to nonspecificity of the method. It may arise because of codetermination of substances that vary in concentration from sample to sample. For example, a chromogenic creatinine method codetermines some other components with creatinine in serum.

The final term in the equation above is the random measurement error term, ε_i. If an infinite number of measurements of a specific sample is performed by the routine method, the random measurement error term ε_i would be zero. The cal-bias and the random-bias$_i$, however, would be unchanged. Thus the average value of an infinite number of measurements would equal the sum of the true value and these bias terms. This average value may be regarded as the target value ($X_{\text{Target}i}$) of the given sample for the routine method:

$$X_{\text{Target}i} = X_{\text{True}i} + \text{Cal-Bias} + \text{Random-Bias}_i$$

As mentioned, the calibration bias represents a systematic error component in relation to the true values measured by a reference measurement procedure. In the context of regression analysis, this systematic error corresponds to the intercept and the slope deviation from unity when a routine method is compared with a reference measurement procedure (outlined in detail later). It is convenient to introduce a modified target value expression ($X'_{\text{Target}i}$) for the routine method to delineate this systematic calibration bias, so that:

$$X'_{\text{Target}i} = X_{\text{True}i} + \text{Cal-Bias}$$

Thus, for a set of samples measured by a routine method, the $X_{\text{Target}i}$ values are distributed around the respective $X'_{\text{Target}i}$ values with an SD, which is called σ_{RB}.

If the assay is a reference method without bias and nonspecificity, the target value and the modified target value equal the true value, that is,

$$X_{\text{Target}i} = X'_{\text{Target}i} = X_{\text{True}i}$$

The error model is outlined in Fig. 2.9.

Calibration Bias and Random Bias

For an individual measurement, the total error is the deviation of x_i from the true value, that is,

$$\text{Total error of } x_i = \text{Cal-Bias} + \text{Random-Bias}_i + \varepsilon_i$$

Estimation of the bias terms requires parallel measurements between the method in question and a reference method. With regard to calibration bias, the possibility of lot-to-lot variation in analytical kit sets should be recognized. Lot-to-lot variation shows up as a calibration bias that changes from lot to lot.

The previous exposition defines the total error in broader terms than is often seen. A traditional total error expression is:

$$\text{Total error} = \text{Bias} + 2\ \text{SD}_A$$

This is often interpreted as the calibration bias plus 2 SD_A. If a one-sided statistical perspective is taken, the expression is modified to Bias + 1.65 SD_A, indicating that 5% of results are located outside the limit. Interpreting the bias as identical with the calibration bias may lead to an underestimation of the total error.

Random bias related to sample-specific interferences may take several forms. It may be a regularly occurring additional random error component, perhaps of the same order of

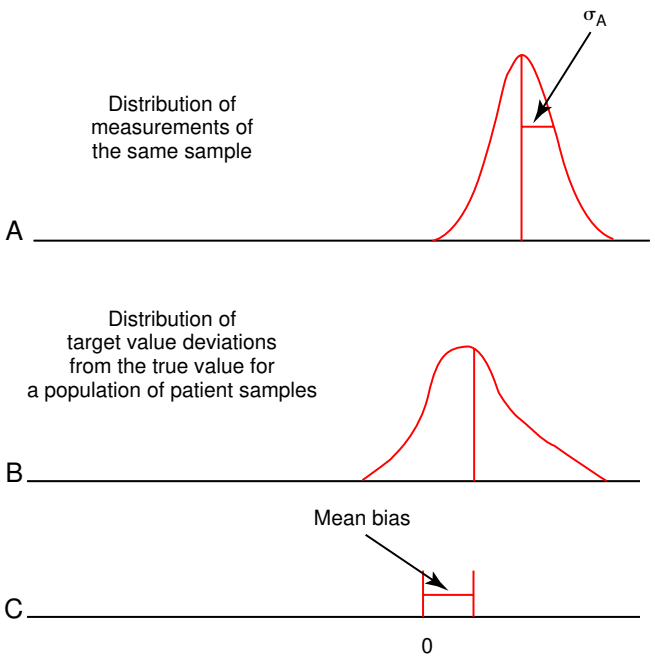


Fig. 2.9 Outline of basic error model for measurements by a routine assay. (A) The distribution of repeated measurements of the *same* sample, representing a normal distribution around the target value ($X_{Target\,i}$) *(vertical line)* of the sample with a dispersion corresponding to the analytical standard deviation, σ_A. (B) Schematic outline of the dispersion of target value deviations from the respective true values for a population of patient samples. A distribution of an arbitrary form is displayed. The standard deviation equals σ_{RB}. The vertical line indicates the mean of the distribution. (C) The distance from zero to the mean of the target value deviations from the true values represents the calibration bias (mean bias = cal-bias) of the assay.

magnitude as the analytical error. In this context, it is natural to quantify the error in the form of an SD or CV. The most straightforward procedure is to carry out a method comparison study based on a set of patient samples in which one of the methods is a reference method, as outlined later. Another form of sample-related random interference is more rarely occurring gross errors, which typically are seen in the context of immunoassays and are related to unexpected antibody interactions. Such an error usually shows up as an outlier in method comparison studies. Outliers should be investigated to identify their cause, which may be an important limitation in using a given assay.

Assay Comparison Data Model

Here we consider the error model described earlier in relation to the method comparison situation. For a given sample measured by two analytical methods, 1 and 2, the following equations apply:

$$x1_i = X1_{Target\,i} + \varepsilon 1_i = X_{True\,i} + \text{Cal-Bias1} + \text{Random-Bias1}_i + \varepsilon 1_i$$

$$x2_i = X2_{Target\,i} + \varepsilon 2_i = X_{True\,i} + \text{Cal-Bias2} + \text{Random-Bias2}_i + \varepsilon 2_i$$

In the following, some typical situations using this general model are described. First, comparison of a routine assay with a reference measurement procedure will be treated. Second, comparison of two routine assays is considered.

Assuming that method 1 is a reference method, the bias components disappear by definition, and the following situation can be described:

$$x1_i = X1_{Target\,i} + \varepsilon 1_i = X_{True\,i} + \varepsilon 1_i$$

$$x2_i = X2_{Target\,i} + \varepsilon 2_i = X_{True\,i} + \text{Cal-Bias2} + \text{Random-Bias2}_i + \varepsilon 2_i$$

The paired differences become

$$(x2_i - x1_i) = \text{Cal-Bias2} + \text{Random-Bias2}_i + (\varepsilon 2_i - \varepsilon 1_i)$$

We thus have an expression consisting of a systematic error term (calibration bias of method 2) and two random terms. The Random-Bias2 term is distributed around Cal-Bias2 according to an undefined distribution. ($\varepsilon 2_i - \varepsilon 1_i$) is a difference between two random measurement errors that are independent and, commonly, Gaussian distributed. However, we remind readers that the SD for analytical methods often depends on the concentration, as mentioned earlier. For analytes with a wide analytical measurement range (e.g., some hormones), both sample-related random interferences and analytical SDs are likely to depend on the measurement concentration, often in a roughly proportional manner. It may then be more useful to evaluate the *relative* differences—($x2_i - x1_i$)/[($x2_i + x1_i$)/2]—and accordingly express mean and random bias and analytical error as proportions.

In the comparison of two routine methods, the paired differences become

$$(x2_i - x1_i) = (\text{Cal-Bias2} - \text{Cal-Bias1}) + (\text{Random-Bias2}_i - \text{Random-Bias1}_i) + (\varepsilon 2_i - \varepsilon 1_i)$$

The expression again consists of a constant term, the difference between the two calibration biases, and two random terms. The first random term is a difference between two random-bias components that may or may not be independent. If the two field methods are based on the same measurement principle, the random bias terms are likely to be correlated. For example, two chromogenic methods for creatinine are likely to be subject to interference from the same chromogenic compounds present in a given serum sample. On the other hand, a chromogenic method and an enzymatic creatinine method are subject to different types of interfering compounds, and the random bias terms may be relatively independent. In the $\varepsilon 2_i - \varepsilon 1_i$ term, the same relationships as described previously are likely to apply. The general form of the expressed differences is the same in the two situations and the same statistical principles apply.

Planning a Method Comparison Study

Several points require attention, including the (1) number of samples necessary, (2) distribution of analyte concentrations (preferably uniform over the analytical measurement range), and (3) representativeness of the samples. To address the

latter point, samples from relevant patient categories should be included, so that possible interference phenomena can be discovered. Practical aspects related to storage and treatment of samples (e.g., container) and possible artifacts induced by storage (e.g., freezing of samples) and addition of anticoagulants should be considered. Comparison of measurements should preferably be undertaken over several days (e.g., at least 5 days) so that the comparison of methods does not become dependent on the performance of the methods in one particular analytical run. The CLSI guideline EP-09-A3, "Method Comparison and Bias Estimation Using Patient Samples," suggests measurement of 40 samples in duplicate by each method when a new method is introduced in the laboratory as a substitute for an established one. Additionally, 100 samples in duplicate is proposed for a vendor of an analytical test system. The EP15 guideline "User Verification of Manufacturer's Claims" suggests a more condensed approach based on a bias or **difference plot** for 20 samples.

Difference (Bland-Altman) Plot

The Bland-Altman plot is usually understood as a plot of the differences against the average results of the methods. Thus the difference plot in this version provides information on the relation between differences and concentration, which is useful in evaluating whether problems exist at certain ranges (e.g., in the high range) caused by nonlinearity of one of the methods. It may also be of interest to observe whether differences tend to increase proportionally with the concentration or whether they are independent of concentration.

The basic version of the difference plot requires plotting of the differences against the average of the measurements. Fig. 2.10 shows the plot for the drug **assay comparison** data. The interval ±2 SD of the differences is often delineated around the mean difference. To assess whether the bias is significantly different from zero, the SE of the mean difference is estimated as the SD divided by the square root of the number of paired measurements ($SE = SD/N^{0.5}$) and tested against zero by a *t*-test ($t = [Mean - 0]/SE$).

Nonparametric limits may also be considered. A constant bias over the analytical measurement range changes the average concentration away from zero. The presence of sample-related random interferences increases the width of the distribution. If the calibration bias depends on the concentration, if the dispersion varies with the concentration, or if both occur, the relations become more complex, and the interval mean ±2 SD of the differences may not fit very well as a 95% interval throughout the analytical measurement range.

The displayed Bland-Altman plot for the drug assay comparison data (see Fig. 2.10) shows a tendency toward increasing scatter with increasing concentration, which is a reflection of increasing random error with concentration. Thus a plot of the relative differences against the average concentration is of relevance (Fig. 2.11).

Regression Analysis

Regression analysis has the advantage that it allows the relation between the target values for the two compared methods

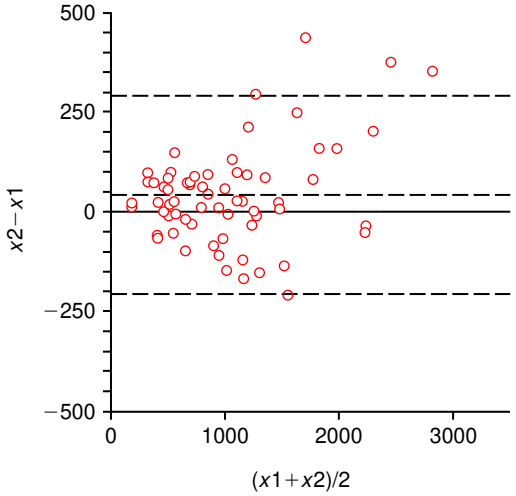


Fig. 2.10 Bland-Altman plot of differences for the drug comparison example. The differences are plotted against the average concentration. The mean difference (42 nmol/L) with ±2 standard deviation of differences is shown *(dashed lines)*.

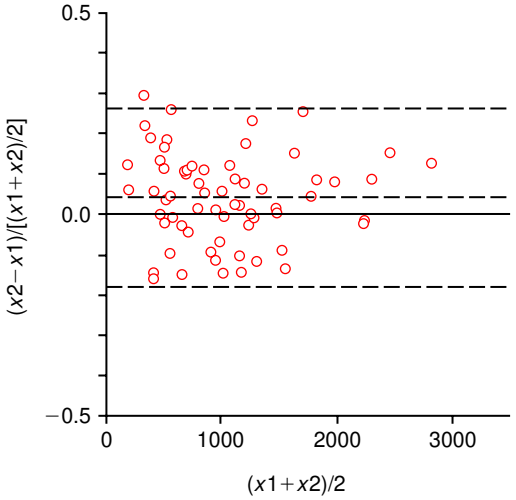


Fig. 2.11 Bland-Altman plot of *relative* differences for the drug comparison example. The differences are plotted against the average concentration. The mean relative difference (0.042) with ±2 standard deviation of relative differences is shown *(dashed lines)*.

to be studied over the full analytical measurement range. In linear regression analysis, it is assumed that the systematic difference between target values can be modeled as a constant systematic difference (intercept deviation from zero) combined with a proportional systematic difference (slope deviation from unity), usually related to a discrepancy with regard to calibration of the methods. In situations when random errors have a constant SD, unweighted regression procedures are used (e.g., **Deming regression** analysis). For cases with SDs that are proportional to the concentration, the weighted Deming regression procedure is preferred.

As outlined previously, we distinguish between the measured value (x_i) and the target value ($X_{Target i}$) of a sample subjected to analysis by a given method. In linear regression analysis, we assume a linear relationship between values devoid of random error of any kind. Thus to operate with a

linear relationship between values without random measurement error and sample-related random bias, we have to introduce modified target values:

$$X1_{\text{Target}i} = X1'_{\text{Target}i} + \text{Random-Bias1}_i >$$

$$X2_{\text{Target}i} = X2'_{\text{Target}i} + \text{Random-Bias2}_i$$

where we now assume a linear relationship between these modified target values:

$$X2'_{\text{Target}i} = \alpha_0 + \beta X1'_{\text{Target}i}$$

In this model, α_0 corresponds to a constant difference with regard to calibration, and $(\beta - 1)$ is a proportional deviation. Thus the systematic error or calibration difference between the measurements corresponds to

$$X2'_{\text{Target}i} - X1'_{\text{Target}i} = \alpha_0 + (\beta - 1)\, X1'_{\text{Target}i}$$

Because of sample-related random interferences and measurement imprecision (of the type that can be described by a Gaussian distribution, e.g., caused by pipetting variability, signal variability), individually measured pairs of values ($x1_i$, $x2_i$) will be scattered around the line expressing the relationship between $X1'_{\text{Target}i}$ and $X2'_{\text{Target}i}$. Fig. 2.12 outlines schematically how the random distribution of $x1$ and $x2$ values occurs around the regression line. We have

$$x1_i = X1_{\text{Target}i} + \varepsilon 1_i = X1'_{\text{Target}i} + \text{Random-Bias1}_i + \varepsilon 1_i$$

$$x2_i = X2_{\text{Target}i} + \varepsilon 2_i = X2'_{\text{Target}i} + \text{Random-Bias2}_i + \varepsilon 2_i$$

The random error components may be expressed as SDs, and generally we can assume that sample-related random bias (SD σ_{RB}) and analytical imprecision (SD σ_{A}) are independent for each analyte, yielding the relations

$$\sigma^2_{ex1} = \sigma^2_{\text{RB1}} + \sigma^2_{\text{A1}}$$

$$\sigma^2_{ex2} = \sigma^2_{\text{RB2}} + \sigma^2_{\text{A2}}$$

σ_{ex1} and σ_{ex2} are the total SDs of the distributions of $x1_i$ and $x2_i$ around their respective modified target values, $X1'_{\text{Target}i}$ and $X2'_{\text{Target}i}$. The sample-related random bias components for methods 1 and 2 may not necessarily be independent. They also may not be Gaussian distributed, contrary to the analytical components. Thus, when a regression procedure is applied, the explicit assumptions to take into account should be considered. In situations without random bias components of any significance, the relationships simplify to

$$\sigma^2_{ex1} = \sigma^2_{\text{A1}}$$

$$\sigma^2_{ex2} = \sigma^2_{\text{A2}}$$

In this situation, it usually can be assumed that the error distributions are Gaussian, and estimates of the analytical SDs may be available from QC data.

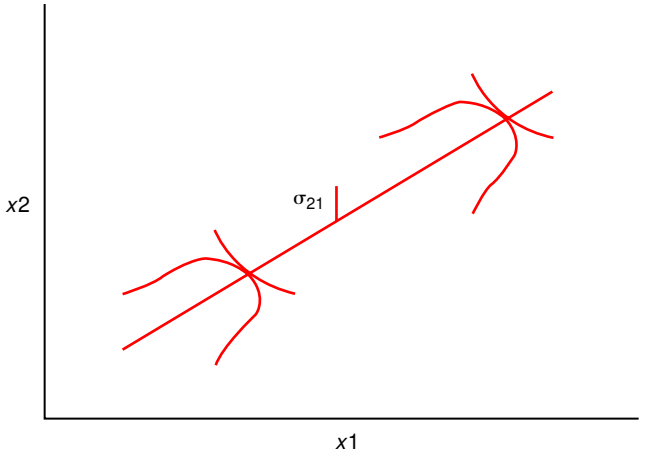


Fig. 2.12 Outline of the relation between $x1$ and $x2$ values measured by two assays subject to random errors with constant standard deviations over the analytical measurement range. A linear relationship between the modified target values ($X1'_{\text{Target}i}$, $X2'_{\text{Target}i}$) is presumed. The $x1_i$ and $x2_i$ values are Gaussian distributed around $X1'_{\text{Target}i}$ and $X2'_{\text{Target}i}$, respectively, as schematically shown. σ_{21} (σ_{yx}) is demarcated.

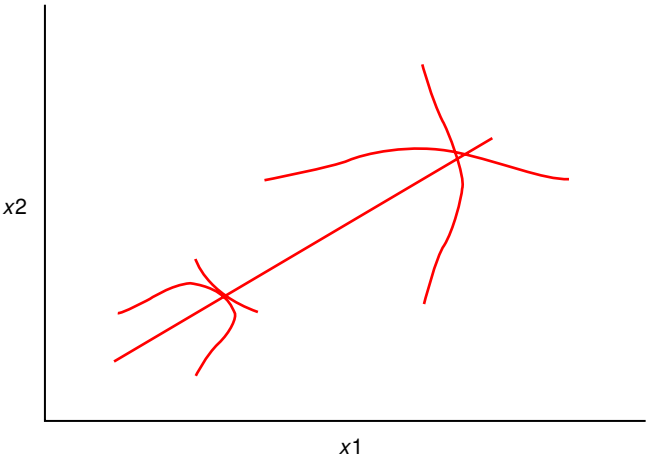


Fig. 2.13 Outline of the relation between $x1$ and $x2$ values measured by two assays subject to proportional random errors. A linear relationship between the modified target values is assumed. The $x1_i$ and $x2_i$ values are Gaussian distributed around $X1'_{\text{Target}i}$ and $X2'_{\text{Target}i}$, respectively, with increasing scatter at higher concentrations, as is shown schematically.

Another methodologic problem concerns the question of whether the dispersion of sample-related random bias and the analytical imprecision are constant or change with the analyte concentration. In cases with a considerable range (i.e., a decade or longer), this phenomenon should also be taken into account when a regression analysis is applied. Fig. 2.13 schematically shows how dispersions may increase proportionally with concentration.

Deming Regression Analysis and Ordinary Least-Squares Regression Analysis (Constant Standard Deviations)

To reliably estimate the relationship between modified target values (i.e., a_0 for α_0 and b for β), a regression procedure taking into account errors in both $x1$ and $x2$ is preferable (i.e., Deming approach) (see Fig. 2.12). Although the **ordinary**

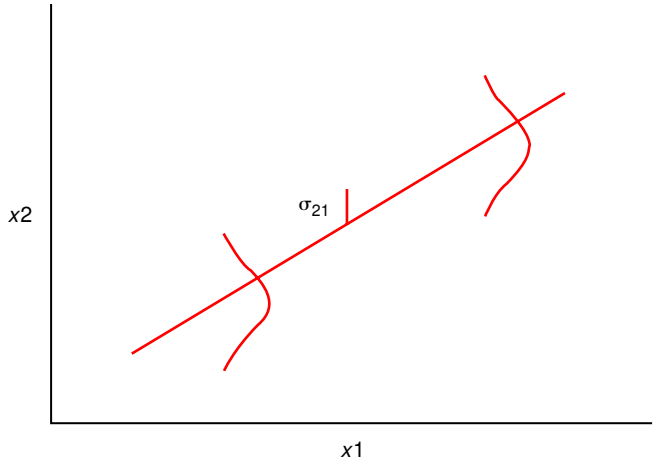


Fig. 2.14 The model assumed in ordinary least-squares regression. The *x*2 values are Gaussian distributed around the line with constant standard deviation over the analytical measurement range. The *x*1 values are assumed to be without random error. σ_{21} (σ_{yx}) is shown.

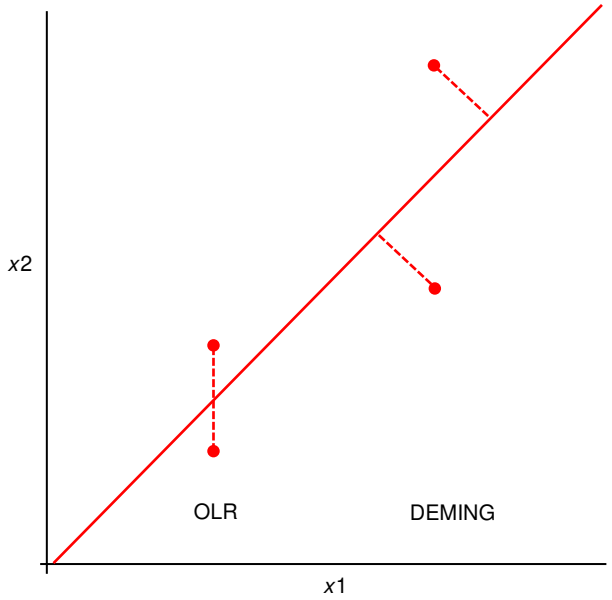


Fig. 2.15 In ordinary least-squares regression *(OLR)*, the sum of squared deviations from the line is minimized in the vertical direction. In Deming regression analysis, the sum of squared deviations is minimized at an angle to the line, depending on the random error ratio. Here the symmetric case is displayed with orthogonal deviations. (From Linnet, K. [1998]. The performance of Deming regression analysis in case of a misspecified analytical error ratio. *Clinical Chemistry, 44,* 1024–1031.)

least-squares regression (OLR) procedure is commonly used in method comparison studies, it does not take errors in *x*1 into account but is based on the assumption that only the *x*2 measurements are subject to random errors (Fig. 2.14). In the Deming procedure, the sum of squared distances from measured sets of values ($x1_i$, $x2_i$) to the regression line is minimized at an angle determined by the ratio between SDs for the random variations of *x*1 and *x*2. It can be proven theoretically that, given Gaussian *error* distributions, this estimation procedure is optimal. It should here be noted that it is the *error* distributions that should be Gaussian, not the dispersion of values over the measurement range. In Fig. 2.15, the symmetric case is illustrated with a regression slope of 1 and equal SDs for the random variations of *x*1 and *x*2, in which case the sum of squared distances is minimized orthogonally in relation to the line.

OLR regression is not recommended except in special situations. In OLR, the sum of squared distances is minimized in the vertical direction to the line (see Fig. 2.15). The neglect of the random error in *x*1 induces a downward biased slope estimate, which depends on the ratio between the SD for the random error in *x*1 and the SD of the *X*1′ target values. Fig. 2.16 shows the bias as a function of the ratio of the random error SD to the SD of the *X*1′ target value dispersion. For a ratio up to 0.1, the bias is less than 1%. At a ratio of 0.33, the bias amounts to 10%. In a given case, the analytical SD (e.g., from QC data) can be divided by the SD of the measured *x*1 values, which approximately equals the SD of *X*1′ target values. As an example, a typical comparison study for two serum sodium methods may be associated with a downward directed slope bias of about 10% (Fig. 2.17).

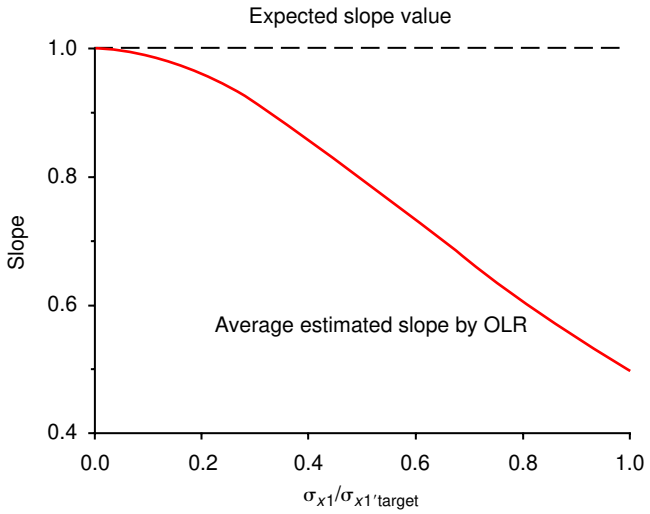


Fig. 2.16 Relations between the true (expected) slope value and the average estimated slope by ordinary least-squares regression *(OLR)*. The bias of the OLR slope estimate increases negatively for increasing ratios of the standard deviation (SD) random error in *x*1 to the SD of the *X*1 target value distribution.

Computation Procedures for Ordinary Least-Squares Regression and Deming Regression

Assuming no errors in *x*1 and a Gaussian error distribution of *x*2 with constant SD throughout the analytical measurement range, OLR is the optimal estimation procedure. Given errors in both *x*1 and *x*2, the Deming approach is the method of choice. It should be noted for these parametric procedures that only the *error* distributions must be Gaussian or normal to ensure the nominal type I errors for associated statistical tests for slope and intercept hold true. The procedures are generally robust toward deviations from normality, but they are sensitive to outliers because of the squaring principle. Finally, the distribution of the *x*1 and *x*2 values over the measurement range does not have to be normal. A uniform distribution over the analytical

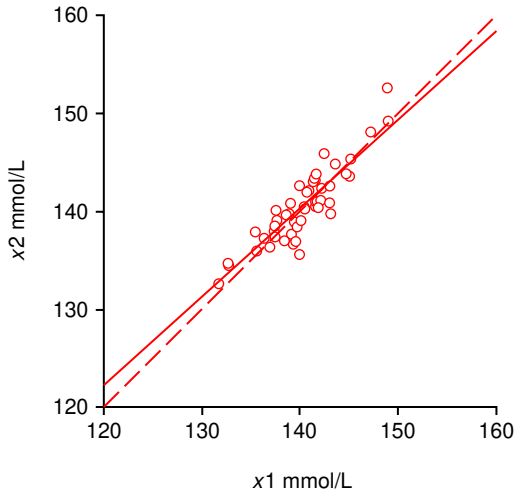


Fig. 2.17 Simulated comparison of two sodium methods. The *solid line* indicates the average estimated ordinary least-squares regression (OLR) line, and the *dotted line* is the identity line. Even though no systematic difference is evident between the two methods, the average OLR line deviates from the identity line corresponding to a downward slope bias of about 10%.

measurement range is generally of advantage, but the distribution in principle may take any form. For both procedures, we may evaluate the SD of the dispersion in the *vertical* direction around the line (commonly denoted $SD_{y\cdot x}$ and here given as SD_{21}). We have

$$SD_{21} = \left[\Sigma(x2_i - X2_{Targetest i})^2/(N-2)\right]^{0.5}$$

Further discussion regarding the interpretation of SD_{21} will be given later.

To compute the slope in Deming regression analysis, the ratio between the SDs of the random errors of $x1$ and $x2$ is necessary, that is,

$$\lambda = \left(\sigma^2_{RB1} + \sigma^2_{A1}\right)/\left(\sigma^2_{RB2} + \sigma^2_{A2}\right)$$

SD_As can be estimated from duplicate sets of measurements as

$$SD^2_{A1} = (1/2N)\left[\Sigma(x1_{2i} - x1_{1i})^2\right]$$

$$SD^2_{A2} = (1/2N)\left[\Sigma(x2_{2i} - x2_{1i})^2\right]$$

If a specific value for λ is not available and the two routine methods that are compared are likely to be associated with random errors of the same order of magnitude, λ can be set to 1. The Deming procedure is generally relatively insensitive to a misspecification of the λ value.

Formulas for computing slope (β), intercept (α_0), and their SEs are available from other sources and are not provided here.

Evaluation of the Random Error Around an Estimated Regression Line

The estimated slope and intercept provide an estimate of the systematic difference or calibration bias between two methods over the analytical measurement range. An estimate of the random error is the dispersion around the line in the vertical direction, which is quantified as $SD_{y\cdot x}$ (here denoted SD_{21}). We have here without sample-related random interferences

$$\sigma^2_{21} = \beta^2\sigma^2_{A1} + \sigma^2_{A2}$$

Thus σ_{21} reflects the random error both in $x1$ (with a rescaling) and in $x2$. Often β is close to unity, and in this case, σ^2_{21} becomes approximately the sum of the individual squared SDs. This relation holds true for both Deming and OLR analyses. Frequently, OLR is applied in situations associated with random measurement error in both $x1$ and $x2$, and in these situations, σ_{21} reflects the errors in both.

The presence of sample-related random interferences in both $x1$ and $x2$ gives the following expression:

$$\sigma^2_{21} = \left(\beta^2\sigma^2_{A1} + \sigma^2_{A2}\right) + \left(\beta^2\sigma^2_{RB1} + \sigma^2_{RB2}\right)$$

Thus the σ_{21} value is influenced by the slope value and the analytical error components σ_{A1} and σ_{A2} (grouped in the first bracket) and σ_{RB1} and σ_{RB2} (grouped in the second bracket). In many cases, the slope is close to unity, in which case we have simple addition of the components. Information on the analytical components is usually available from duplicate sets of measurements or from QC data. On this basis, the combined random bias term in the second bracket can be derived by subtracting the analytical components from σ_{21}. Overall, it can be judged whether the total random error is acceptable or not. The systematic difference can be adjusted for relatively easily by rescaling one of the sets of measurements. However, if the random error term is very large, such a rescaling does not ensure equivalency of measurements with regard to individual samples.

Assessment of Outliers

The distance from a suspected outlier to the line is recorded in SD units, and the outlier is rejected if the distance exceeds a predetermined limit (e.g., 3 or 4 SD units). In the case of OLR, the SD unit equals SD_{21}, and the vertical distance is considered. For Deming regression analysis, the unit is the SD of the deviation of the points from the line at an angle determined by the error variance ratio λ. A plot of these deviations, a so-called residuals plot, conveniently illustrates the occurrence of outliers. Fig. 2.18A illustrates an example of Deming regression analysis with occurrence of an outlier and the associated residuals plot (see Fig. 2.18B), which clearly shows the outlier pattern. In this example, the residuals plot was standardized to unit SD. Use of an outlier limit of 4 SD units in this example led to rejection of the outlier, and a reanalysis was undertaken. In this example, rejection of the outlier changed the slope from 1.14 to 1.03. With regard to outliers, the reason for their presence should be investigated as a method limitation (e.g., possibly a nonspecificity for the analyte).

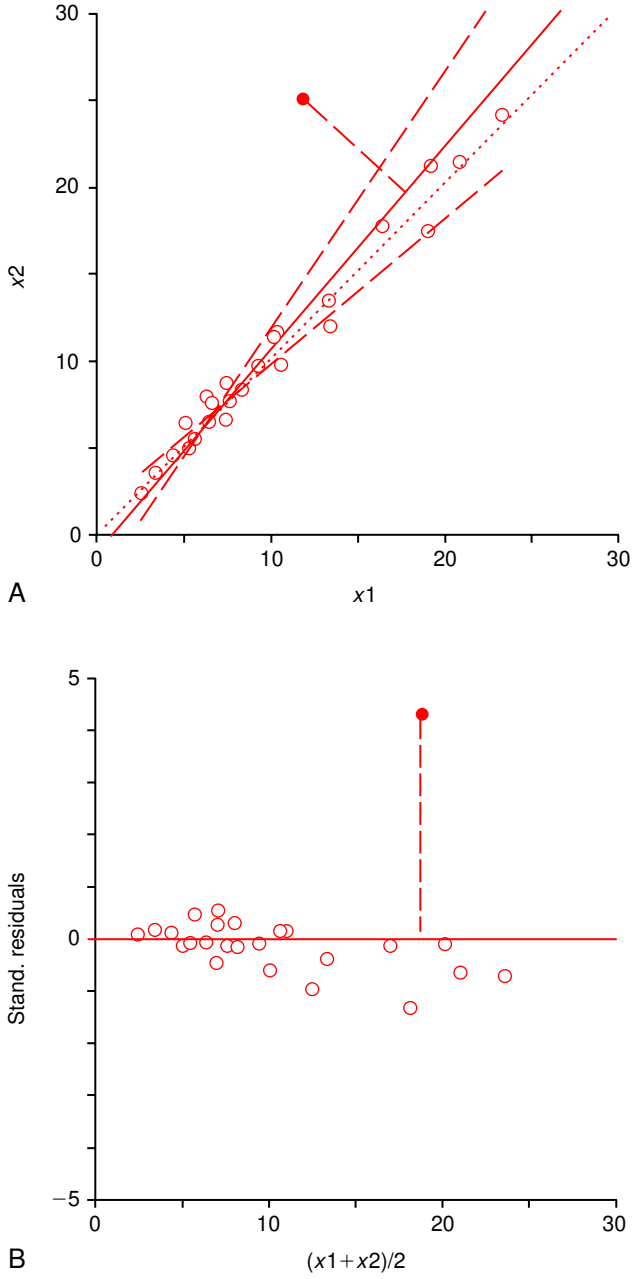


Fig. 2.18 (A) A scatter plot with the Deming regression line *(solid line)* with an outlier *(filled point)*. The *dotted straight line* is the diagonal, and the *curved dashed lines* demarcate the 95% confidence region. (B) Standardized residuals plot with indication of the outlier.

Correlation Coefficient

The ordinary **correlation coefficient**, ϱ, also called the Pearson product moment correlation coefficient, is estimated as r from sums of squared deviations for $x1$ and $x2$ values as follows:

$$r = p/(uq)^{0.5}$$

where

$$p = \Sigma\,(x1_i - x1_m)\,(x2_i - x2_m)$$

$$u = \Sigma(x1_i - x1_m)^2 \text{ and } q = \Sigma(x2_i - x2_m)^2$$

and

$$x1_m = \Sigma x1_i/N \text{ and } x2_m = \Sigma x2_i/N$$

The correlation coefficient r is a *relative* indicator of the amount of dispersion around the regression line. If the numeric interval of values is short, r tends to be low and vice versa for a long range of values. For example, consider simulated examples, where the random errors of $x1$ and $x2$ are the same but the width of the distributions of measured values differs (Fig. 2.19A and B). In A, the target values are uniformly distributed over the range 1 to 3, and in B, the range is 1 to 6. The random error SD is presumed constant, and it is set to 0.15 for both $x1$ and $x2$, corresponding to a CV of 5% at the value 3. Given sets of 50 paired measurements, the correlation coefficient is 0.93 in case *A* and 0.99 in case *B*. Furthermore, a single point located outside the range of the rest of the observations exerts a strong influence, resulting in a value of 0.97 (see Fig. 2.19C).

Although σ_{21} is the relevant measure for random error in method comparison studies, ϱ is still incorrectly used as a supposed measure of agreement between two methods. A systematic difference due to a difference with regard to calibration is not expressed through ϱ but solely in the form of an intercept (α_0) deviation from zero or a slope (β) deviation from unity.

Regression Analysis in Cases of Proportional Random Error

For analytes with extended ranges (e.g., 1 or several decades), the SD_A is seldom constant. Rather, a proportional relationship may apply. This may also be true for the random bias components. In this situation, the regression procedures described previously may still be used, but they are not optimal because the SEs of slope and intercept become larger than is the case when a weighted form of regression analysis is applied. Given a proportional relationship, a weighted procedure assigns larger weights to observations in the low range; low-range observations are more precise than measurements at higher concentrations that are subject to larger random errors. More specifically, weights are applied in the computations that are inversely proportional to the squared SDs (variances) that express the random error. In the weighted modification of the Deming procedure, distances from ($x1_i$, $x2_i$) to the line are inversely weighted according to the squared SDs at a given concentration (Fig. 2.20).

Testing for Linearity

Splitting of the systematic error into a constant and a proportional component depends on the assumption of linearity. A convenient test is a runs test, which in principle assesses whether negative and positive deviations from the points to the line are randomly distributed. The term *run* here relates to a sequence of deviations with the same sign. Consider, for example, the situation with a downward trend of $x2$ values at the upper end of the analytical measurement range (Fig. 2.21A). The SDs from the line (i.e.,

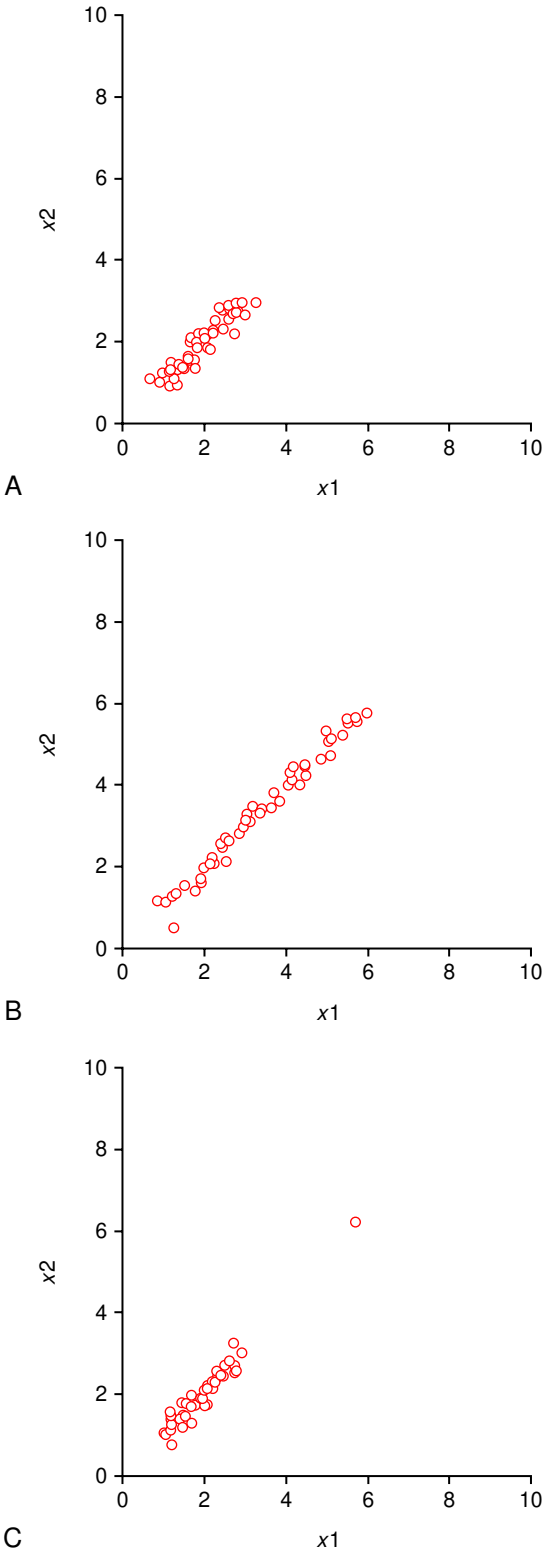


Fig. 2.19 Scatter plots illustrating the effect of the range on the value of the correlation coefficient ϱ. (A) Target values are uniformly distributed over the range 1 to 3 with random errors of both *x*1 and *x*2 corresponding to a standard deviation (SD) of 5% of the target value at 3 (constant error SDs). (B) The range is extended to 1 to 6 with the same random error levels. The correlation coefficient equals 0.93 in A and 0.99 in B. (C) The effect of a single aberrant point is shown. Forty-nine of the target values are distributed over the range 1 to 3, with a single point at 6. The correlation coefficient is 0.97.

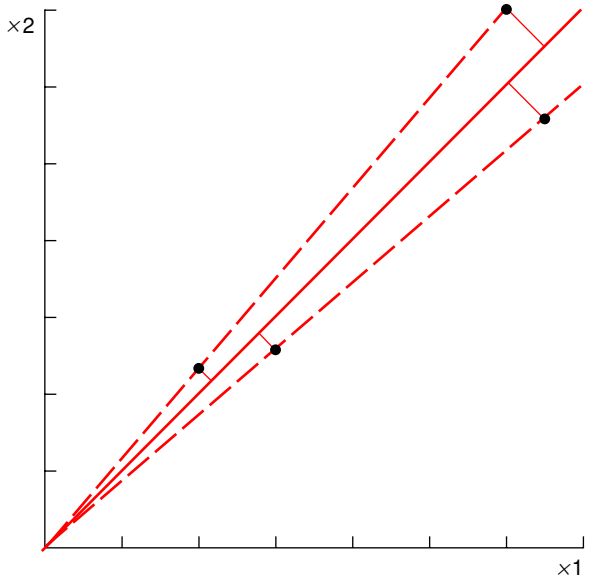


Fig. 2.20 Distances from data points to the line in weighted Deming regression assuming proportional random errors in *x*1 and *x*2. The symmetric case is illustrated with equal random errors and a slope of unity yielding orthogonal projections onto the line. (Modified from Linnet, K. [1999]. Necessary sample size for method comparison studies based on regression analysis. *Clinical Chemistry, 45,* 882–894. Used with permission.)

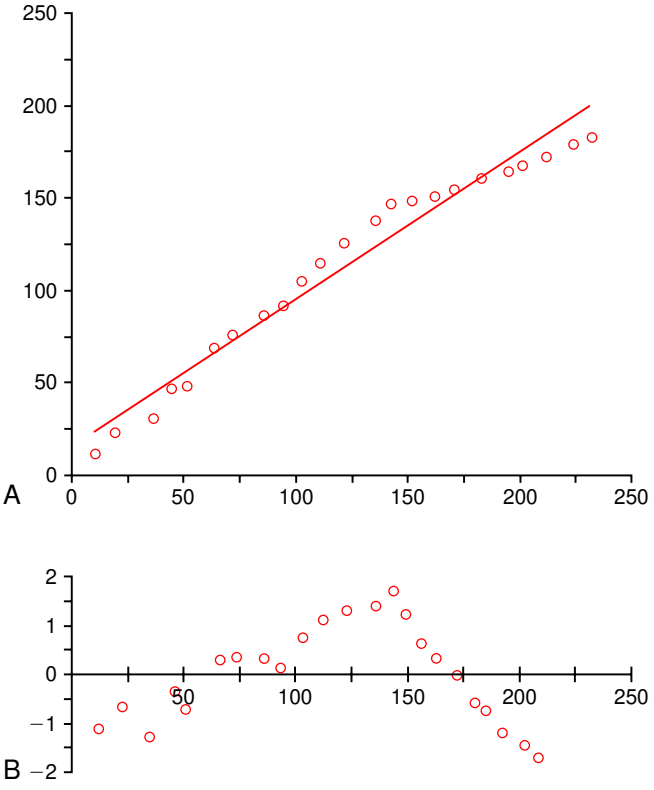


Fig. 2.21 (A) Scatter plot showing an example of nonlinearity in the form of downward-deviating *x*2 values at the upper part of the range. (B) Plot of residuals showing the effects of nonlinearity. At the upper end of the analytical measurement range, a sequence (run) of negative residuals is present.

the residuals) will tend to be negative in this area instead of being randomly distributed above and below the line (see Fig. 2.21B).

Nonparametric Regression Analysis (Passing-Bablok Procedure)

The slope and the intercept may be estimated by a nonparametric procedure, which is robust to outliers and requires no assumptions of Gaussian error distributions. The method takes measurement errors for both $x1$ and $x2$ into account, but it presumes that the ratio between random errors is related to the slope in a fixed manner:

$$\lambda = \left(SD^2_{RB1} + SD^2_{A1}\right) / \left(SD^2_{RB2} + SD^2_{A2}\right) = 1/\beta^2$$

Otherwise, a biased slope estimate is obtained. The procedure may be applied both in situations with random errors with constant SDs and in cases with proportional SDs. The method is not as efficient as the corresponding parametric procedures. Slope and intercept with CIs are provided, together with Spearman's rank correlation coefficient.

Interpretation of Systematic Differences Between Methods Obtained on the Basis of Regression Analysis

A systematic difference between two methods is identified if the estimated intercept differs significantly from zero or if the slope deviates significantly from 1. This is decided on the basis of t-tests:

$$t = (a_0 - 0)/SE(a_0)$$

$$t = (b - 1)/SE(b)$$

The t-tests can be supplemented with 95% CIs.

SE(a_0) and SE(b) are the SEs of the estimated intercept a_0 and the slope b, respectively. SEs can be derived by a computerized resampling principle called *the jackknife procedure*, which in practice can be carried out using appropriate software. Having estimated a_0 and b, we have the estimate of the systematic difference between the methods, D_c, at a selected concentration, $X1'_{Targetc}$:

$$D_c = X2'_{Targetestc} - X1'_{Targetc} = a_0 + (b - 1)\, X1'_{Targetc}$$

$X2'_{Targetestc}$ is the estimated $X2'$ target value at $X1'_c$. Note that D_c refers to the *systematic difference* (i.e., the difference between modified target values corresponding to a calibration difference). The SE of D_c can be derived by the jackknife procedure using a software program. By evaluating the SE throughout the analytical measurement range, a confidence region for the estimated line can be displayed. If method comparison is performed to assess the calibration to a reference measurement procedure, correction of a significant systematic difference $Delta_c$ will often be performed by recalibration [$x2_{rec} = (x1 - a_0)/b$]. The associated standard uncertainty is the SE of $Delta_c$.

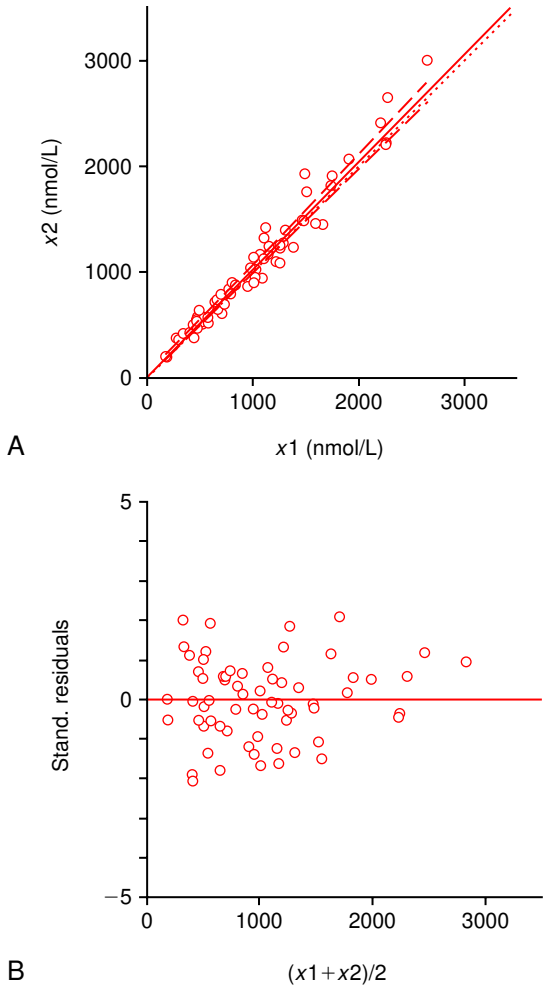


Fig. 2.22 An example of weighted Deming regression analysis for the comparison of drug assays. (A) The *solid line* is the estimated weighted Deming regression line, the *dashed curves* indicate the 95% confidence region, and the *dotted line* is the line of identity. (B) A plot of residuals standardized to unit standard deviation. The homogeneous scatter supports the assumed proportional error model and the assumption of linearity.

Example of Application of Regression Analysis (Weighted Deming Analysis)

Application of weighted Deming regression analysis may be illustrated by the comparison of drug assays example [N = 65 ($x1$, $x2$) single measurements]. As outlined in the section on the Bland-Altman plot (see Fig. 2.10), in this example the random error of the differences increases with the concentration, suggesting that the weighted form of Deming regression analysis is appropriate. Fig. 2.22 shows (A) the estimated regression line with 95% confidence bands and (B) a plot of normalized residuals. The nearly homogeneous scatter in the residuals plot supports the assumed proportional random error model and the assumption of linearity. The slope estimate (1.014) is not significantly different from 1 (95% CI: 0.97 to 1.06), and the intercept is not significantly different from zero (95% CI: −6.7 to 47.4) (Table 2.2). A runs test for linearity does not contradict the assumption of linearity. The amount of random error is quantified in the form

TABLE 2.2 Results of Weighted Deming Regression Analysis for the Comparison of Drug Assays Example, *n* = 65 Single (*x*1, *x*2) Measurements

	Estimate	SE	95% CI
Slope *(b)*	1.014	0.022	0.97–1.06
Intercept (a_0)	20.3	13.5	−6.7 to 47.4
Weighted correlation coefficient	0.98		
SD_{21} proportionality factor	0.11		
Runs test for linearity	NS		
$Delta_c = X_2 - X_1$ at $X_c = 300$	24.6	9.5	5.72–43.6
$Delta_c = X_2 - X_1$ at $X_c = 2000$	48.9	34.2	−19.3 to 117

CI, Confidence interval; *NS*, not significant; *SD*, standard deviation; *SE*, standard error.

of the SD_{21} proportionality factor equal to 0.11, or 11%. In the present example, with a slope close to unity and two routine methods with assumed random errors of about the same magnitude, we divide the random error by the square root of 2 and get $CV_{x1} = CV_{x2} = 7.8\%$. QC data in the laboratory have provided CV_As of 6.1% and 7.2% for methods 1 and 2, respectively. Thus in this example, the random error may be attributed largely to analytical error. The assay principle for both methods is high performance liquid chromatography (HPLC), which generally is a rather specific measurement principle; considerable random bias effects are not expected in this case.

In Table 2.2, estimated systematic differences at the limits of the therapeutic interval (300 and 2000 nmol/L) are displayed (24.6 and 48.9 nmol/L, respectively). This corresponds to percentage values of 8.2% and 2.4%, respectively. Estimated SEs by the jackknife procedure yield the 95% CIs, as shown in the table. At the low concentration, the difference is significant (95% CI: 5.7 to 44 nmol/L; does not include zero), which is not the case at the high level (95% CI: −19 to 117 nmol/L). Even though the intercept and slope estimates separately are not significantly different from the null hypothesis values of 0 and 1, respectively, the combined difference $Delta_c$ is significant at low concentrations in this example. If the difference is considered of medical importance and both methods are to be used simultaneously in the laboratory, recalibration of one of the methods might be considered.

MONITORING SERIAL RESULTS

An important aspect of clinical chemistry is monitoring of disease or treatment (e.g., tumor markers in cases of cancer, drug concentrations in cases of therapeutic drug monitoring). To assess changes in a rational way, various imprecision components have to be taken into account. Biologic within-subject variation (SD_I) and preanalytical (SD_{PA}) and analytical variation (SD_A) all have to be recognized. Assuming that preanalytical variation is already included in the estimated within-subject variation SD, a total SD (SD_T) can be estimated as follows:

$$SD_T^2 = SD_{WithinB}^2 + SD_A^2$$

The limit for statistically significant changes then is [*fx*2], where *k* depends on the desired probability level. Considering a two-sided 5% level, *k* is 1.96. The corresponding one-sided factor is 1.65. If a higher probability level is desired, *k* should be increased.

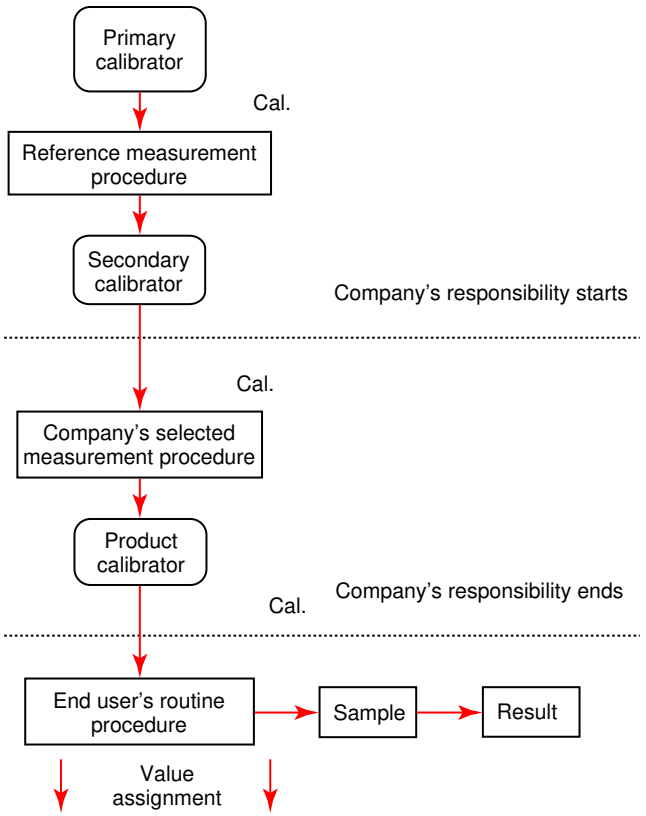


Fig. 2.23 The calibration hierarchy from a reference measurement procedure to a routine assay. The uncertainty increases from top to bottom. *Cal.*, Calibration.

TRACEABILITY

To ensure reasonable agreement between measurements of routine methods, the concept of traceability comes into focus. (See also Chapter 7 for further detail.) **Traceability** is based on an unbroken chain of comparisons of measurements leading to a known reference value (Fig. 2.23). For well-established analytes, a hierarchy of methods exists with *a reference measurement procedure* at the top, *selected measurement procedures* at an intermediate level, and finally *routine measurement procedures* at the bottom. A **reference measurement procedure** is a fully understood procedure of highest analytical quality containing a complete uncertainty budget given in Système Internationale (SI) units. Reference procedures are used to measure the analyte concentration in *secondary reference materials*, which typically have the same matrix as samples that are to be measured by routine procedures (e.g., human serum). Secondary reference materials are usually of high

analytical quality, and certified secondary reference materials must be validated for **commutability** with clinical samples if they are intended for use as trueness controls for routine methods. Otherwise, their use is restricted to selected measurement procedures for which they are intended. The certificate of analysis should state the methods for which the secondary reference materials have been validated to be commutable with clinical samples. When no information is given for commutability, it must be assumed that the reference material is not commutable with clinical samples, and the user has the responsibility to validate commutability for the methods of interest.

Using cortisol as an example, the primary reference material is crystalline cortisol with a chemical analysis for impurities (National Institute of Standards and Technology [NIST] standard reference material 921, cortisol [hydrocortisone]). A primary calibrator is then a cortisol preparation with a stated mass fraction (purity) (e.g., 0.998 and a 95% CI of ±0.001). The reference measurement procedure is an isotope-dilution gas chromatography–mass spectrometry method that is calibrated with the primary calibrator. A panel of individual frozen serum samples that have values assigned by the primary reference measurement procedure is available from the Institute for Reference Materials and Measurements (IRMM) as secondary reference materials (European Reference Material [ERM]-DA451/International Federation of Clinical Chemistry and Laboratory Medicine [IFCC]). A manufacturer's *selected measurement* procedure is calibrated with the secondary reference materials and is used for measurement of the quantity in the manufacturer's *product calibrator*, which is the calibrator used for the routine method in clinical laboratories.

In case a reference measurement procedure exists for an analyte (measurand), comparable results among measurement procedures can be achieved as described earlier, so-called standardization. When reference measurement procedures are not available, so-called harmonization refers to the process of establishing comparable results among measurement procedures for the given analyte. Harmonization is typically based on distribution among laboratories of commutable secondary reference materials with arbitrarily set target values (see Chapter 7).

Uncertainty Concept

According to the ISO's "Guide to the Expression of Uncertainty in Measurement" (GUM), **uncertainty** is formally defined as "a parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand." In practice, this means that the uncertainty is given as an interval around a reported laboratory result that specifies the location of the true value with a given probability (e.g., 95%). In general, the uncertainty of a result, which is traceable to a particular reference, is the uncertainty of that reference together with the overall uncertainty of the traceability chain.

In the outline of the uncertainty concept, it is assumed that any known systematic error components of a measurement method have been corrected, and the specified uncertainty includes uncertainty associated with correction of the systematic error(s). A distinction between type A and B uncertainties is made. Type A uncertainties are frequency-based estimates of SDs (e.g., an SD of the imprecision). Type B uncertainties are uncertainty components for which frequency-based SDs are not available. Instead, uncertainty is estimated by other approaches or by the opinion of experts. Finally, the total uncertainty is derived from a combination of all sources of uncertainty and can be expressed as a *standard uncertainty* (u_{st}), which is equivalent to an SD. By multiplication of a standard uncertainty with a *coverage factor (k)*, the uncertainty corresponding to a specified probability level is derived, e.g. multiplication with a coverage factor of 2 yields a probability level of ≈95%, given a Gaussian distribution. When the total uncertainty of an analytical result obtained by a routine method is considered (u_{st}), preanalytical variation (u_{PAst}), method imprecision (u_{Ast}), sample-related random interferences (u_{RBst}), and uncertainty related to calibration and bias corrections (traceability) (u_{Tracst}) should be taken into account. In expressing the uncertainty components as standard uncertainties, we have:

$$u_{st} = \left[u_{PAst}^2 + u_{Ast}^2 + u_{RBst}^2 + u_{Tracst}^2\right]^{0.5}$$

In principle, uncertainty can be judged *directly* from measurement comparisons ("top down") or *indirectly* from an analysis of individual error sources according to the law of error propagation ("error budget," "bottom up").

Example of Direct Assessment of Uncertainty on the Basis of Measurements of a Commutable Certified Reference Material

Suppose a CRM is available that was validated to be *commutable* with patient samples for a given routine method with a specified value 10.0 mmol/L and a standard uncertainty of 0.2 mmol/L. Ten repeated measurements in independent runs give a mean value of 10.3 mmol/L with SD 0.5 mmol/L. The SE of the mean is then [$fx3$]. The mean is not significantly different from the assigned value [$t = (10.3 - 10.0)/(0.2^2 + 0.16^2)^{0.5} = 1.17$]. The total standard uncertainty with regard to traceability is then $u_{Tracst} = (0.16^2 + 0.2^2)^{0.5} = 0.26$ mmol/L. If the bias had been significant, a correction to the method could have been carried out, and the standard uncertainty would then be the same at the given concentration. Thus measurements of the CRM provide an estimate of the uncertainty related to traceability, *given the assumption of commutability with patient samples*. The other components have to be estimated separately. Concerning method imprecision, long-term imprecision (e.g., observed from QC measurements) should be used rather than the short-term SD observed for CRM material. Here we suppose that the long-term SD_A is 0.8 mmol/L. Data on preanalytical variation can be obtained by sampling in duplicates from a series of patients or can be a matter of judgment (type B uncertainty) based on literature data or data on similar analytes. We here suppose that SD_{PA} equals half the analytical SD (i.e., 0.4 mmol/L). Finally, we lack data on a possible sample-related random bias component,

which we may choose to ignore in the present example. The standard uncertainty of the results then becomes

$$u_{st} = \left[u^2_{PAst} + u^2_{Ast} + u^2_{RBst} + u^2_{Tracst}\right]^{0.5}$$
$$= \left(0.4^2 + 0.8^2 + 0.26^2\right)^{0.5}$$
$$= 0.93\ (\text{mmol/L})$$

In this case, the major uncertainty component is the long-term imprecision in the laboratory. To attain a reasonably precise uncertainty estimate, estimated SDs should be based on an appropriate number of repetitions. In the subsection on method precision, it can be seen that $N = 30$ repetitions provides SD estimates with 95% CIs extending from about 20% below to 35% above an estimated value (see Fig. 2.7), which may be regarded as reasonable.

Indirect Evaluation of Uncertainty by Quantification of Individual Error Source Components

On the basis of a detailed quantitative model of the analytical procedure, the standard approach is to assess the standard uncertainties associated with individual input parameters and combine them according to the law of propagation of uncertainties. The relationship between the combined standard uncertainty $u_c(y)$ of a value y and the uncertainty of the *independent* parameters $x_1, x_2, \ldots x_n$, on which it depends, is

$$u_c\left[y(x_1, x_2, \ldots)\right] = \left[\Sigma c_i^2 u(x_1)^2\right]^{0.5}$$

where c_i is a sensitivity coefficient (the partial differential of y with respect to x_i). These sensitivity coefficients indicate how the value of y varies with changes in the input parameter x_i. If the variables are not independent, the relationship becomes

$$u_c\left[y(x_1, x_2, \ldots)\right] = \left[\Sigma c_i^2 u(x_1)^2 + \Sigma c_i c_k u(x_i, x_k)^2\right]^{0.5}$$

where $u(x_i, x_k)$ is the covariance between x_i and x_k, and c_i and c_k are the sensitivity coefficients. The covariance is related to the correlation coefficient ϱ_{ik} by

$$u(x_i, x_k) = u(x_i)\, u(x_k)\, \rho_{ik}$$

This is a complex relationship that usually will be difficult to evaluate in practice. In many situations, however, the contributing factors are independent, thus simplifying the picture. Below, some simple examples of combined expressions are shown. The rules are presented in the form of combining SDs or CVs given *independent* input components.

$$q = x + y \quad SD(q) = \left[SD(x)^2 + SD(y)^2\right]^{0.5}$$

$$q = x - y \quad SD(q) = \left[SD(x)^2 + SD(y)^2\right]^{0.5}$$

$$q = ax \quad SD(q) = aSD(x) \quad \text{and} \quad CV(q) = CV(x)$$

$$q = x^p \quad CV(q) = p\ CV(x)$$

$$q = xy \quad CV(q) = \left[CV(x)^2 + CV(y)^2\right]^{0.5}$$

$$q = x/y \quad CV(q) = \left[CV(x)^2 + CV(y)^2\right]^{0.5}$$

The formulas shown may be used, for example, to calculate the combined uncertainty of a calibrator solution from the uncertainties of the reference compound, the weighting, and dilution steps. In some situations, a Monte Carlo simulation model of a complex analytical method may be established to estimate the combined uncertainty of the method on the basis of input uncertainties.

POINTS TO REMEMBER

- For well-established analytes, a hierarchy of methods exists with *a reference measurement procedure* at the top, *selected measurement procedures* at an intermediate level, and finally *routine measurement procedures* at the bottom.
- The uncertainty is given as an interval around a reported laboratory result that specifies the location of the true value with a given probability (e.g., 95%).
- The uncertainty of a result, which is traceable to a particular reference, is the uncertainty of that reference together with the overall uncertainty of the traceability chain.
- The uncertainty can be judged *directly* from measurement comparisons ("top down") or *indirectly* from an analysis of individual error sources according to the law of error propagation ("error budget," "bottom up").

DIAGNOSTIC ACCURACY OF LABORATORY TESTS

We here consider the basic steps for evaluation of the clinical accuracy of laboratory tests. In diagnostic accuracy studies, the measurements or results of one (or more) laboratory test under evaluation (i.e., the so-called index test) are compared with the results of a reference standard or method. This reference is the best prevailing test or strategy that is used to establish the presence or absence of the disease of interest (i.e., the so-called target disease that is to be detected or excluded by the index tests). This reference standard is conducted and its results interpreted as blindly for and independently from the index test(s) results as possible. Test accuracy studies show the concordance in results of the index test(s) with the presence or absence of disease as defined by the reference standard results. These studies provide information regarding the frequency of types of errors (i.e., false positive and negative test results) by the index test in relation to the reference standard.

Diagnostic Accuracy, Sensitivity, and Specificity of a Test in Isolation

In a diagnostic accuracy study the results of the index test are compared with those of a reference test in the same

	Disease status	
Test result	Diseased	Nondiseased
Positive	TP	FP
Negative	FN	TN

Fig. 2.24 The basic 2-by-2 table for estimating the diagnostic accuracy of a dichotomized quantitative test result. Positive test results are divided into true positives *(TPs)* and false positives *(FPs)* and negative results into true negatives *(TNs)* and false negatives *(FNs)*. (From Linnet, K., Bossuyt, P. M., Moons, K. G., & Reitsma, J. B. [2012]. Quantifying the accuracy of a diagnostic test or marker. *Clinical Chemistry, 58,* 1292–1301.)

individuals, all of whom are suspected to have the target disease (the suspected disease cohort design). The simplest situation is a comparison of a single index test, with only two result categories (i.e., a dichotomous or binary index test) to a reference standard (i.e., a single-test accuracy study). The ideal dichotomous index test correctly identifies all individuals as diseased or nondiseased with an error rate of zero. A zero error rate is only possible when there is no overlap between index test results in the diseased and nondiseased individuals. However, when there is overlap in index test results, some individuals are classified wrongly as shown below in an example concerning the diagnosis of deep venous thrombosis (DVT) using a D-dimer index test. When using a quantitative (continuous) index test to classify individuals as diseased or nondiseased, a cutoff value needs to be chosen to estimate these error rates. This results in a so-called dichotomized index test.

Values of the dichotomous or dichotomized index test that exceed the cutoff in individuals having the target disease are classified as true positives (TP) (Fig. 2.24). Similarly, index test results lower than the cutoff in nondiseased individuals are true negatives (TN). Accordingly, index test results below the cutoff in truly diseased subjects are false negative (FN), and index test results exceeding the cutoff in truly nondiseased subjects are false positive (FP). Based on the frequencies of FN and FP results, an overall error rate or non-error rate can be derived. The overall diagnostic accuracy of an index test is then defined as the fraction of true classifications out of all classifications:

$$\text{Diagnostic accuracy} = (TN + TP) / (TN + TP + FP + FN)$$

This is an overall non-error rate that can be subdivided into the non-error rate of the nondiseased individuals, which is the specificity of the test and the non-error rate of diseased individuals which is the sensitivity of the test

$$\text{Specificity} = TN / (TN + FP)$$

$$\text{Sensitivity} = TP / (TP + FN)$$

Whereas a very specific test provides negative results for all or almost all subjects who are free of the target disease, a very sensitive test detects all or almost all diseased subjects. To assess the (im)precision of these estimates, CIs should be specified as described under *Categorical Variables*. Table 2.3 displays the widths of the 95% CIs at various sample sizes of

TABLE 2.3 Relationship Between Sample Size and 95% Confidence Intervals of a Proportion (e.g., a Sensitivity or Specificity): Selected Examples of Proportions of 0.05 and 0.80

Sample Size	95% CI of a Proportion of 0.05	95% CI of a Proportion of 0.80
20	0.00–0.25	0.56–0.94
60	0.01–0.14	0.68–0.90
100	0.02–0.11	0.71–0.87
500	0.03–0.07	0.76–0.83
1000	0.04–0.07	0.77–0.82

CI, Confidence interval.

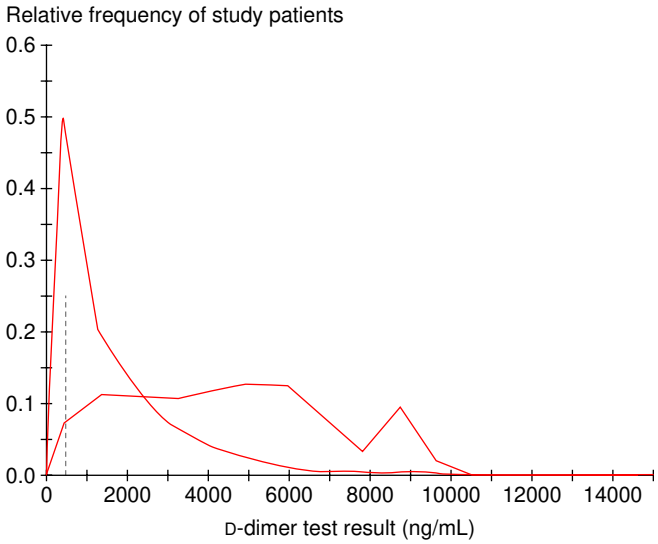


Fig. 2.25 Distribution of the quantitative D-dimer values for deep venous thrombosis (DVT) and non-DVT subjects in the example study. *Light blue line,* non-DVT; *blue line,* DVT. The *dashed line* indicates the commonly used cutoff value of 500 μg/L. (From Linnet, K., Bossuyt, P. M., Moons, K. G., & Reitsma, J. B. [2012]. Quantifying the accuracy of a diagnostic test or marker. *Clinical Chemistry, 58,* 1292–1301.)

20 to 1000 for two selected proportions. The specificity and sensitivity of two tests applied in the same study subjects can be statistically compared using the McNemar's test.

Clinical Example: Accuracy of D-Dimer Test in Diagnosis of Deep Venous Thrombosis

We illustrate the concepts using some of the empirical data of a previously published study in primary care patients suspected of having DVT, the target disease (Fig. 2.25).

The study consisted of 2086 patients suspected of DVT, where DVT was defined as present in patients manifesting at least one of the following symptoms or signs: presence of swelling, redness, or pain in the leg. All patients were given a standardized diagnostic workup, including medical history; clinical examination; and testing for D-dimer, the (quantitative) index test. The reference procedure consisted of repeated compression ultrasonography tests and was performed in all patients, blinded to and independent of the index test results. A total of 416 (20%) of the 2086 included patients had DVT.

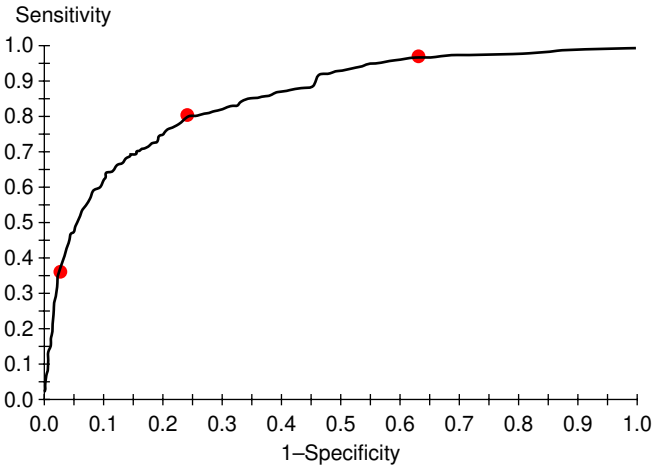


Fig. 2.26 Receiver operating characteristic curve of the D-dimer assay result for diagnosis of deep venous thrombosis in our example study. The *blue markers* correspond to various cutoff choices (from left to right, 5435 µg/L, 2133 µg/L, and 500 µg/L). (From Linnet, K., Bossuyt, P. M., Moons, K. G., & Reitsma, J. B. [2012]. Quantifying the accuracy of a diagnostic test or marker. *Clinical Chemistry, 58,* 1292–1301.)

Applying a commonly used cutoff of 500 µg/L or greater for the (originally) quantitative D-dimer assay (dashed line in Fig. 2.25), the sensitivity was 0.97 (i.e., 3% of the subjects with DVT had a value <500 µg/L). The specificity was only 0.37. The resulting overall diagnostic accuracy was 0.50. Whereas the test displayed good sensitivity at this threshold, detecting all but 3% of those having DVT, its specificity at this test threshold was relatively low, resulting in many FP results. The SEs were 0.012 for the specificity and 0.008 for the sensitivity, resulting in CIs of 0.356 to 0.402 and 0.955 to 0.987, respectively.

Receiver Operating Characteristic Curves

For a quantitative index test, the specificity and sensitivity depend on the selected cutoff point. A plot of the sensitivity and specificity pairs for all possible cutoff values over the measurement range provides the so-called **receiver operating characteristic (ROC) curve** (Fig. 2.26) Usually, sensitivity *(y)* is plotted against (1 – specificity) *(x)* at each possible cutoff value. The better the performance of the test, the higher the ROC curve is located in the left, upper region of the plot. With use of the ROC curve, an appropriate combination of specificity and sensitivity may be chosen, and the corresponding cutoff then selected.

An area under the ROC curve (i.e., the ROC area or so-called concordance or c-index) can be assessed by either parametric or nonparametric statistics. Given an SE of the ROC area or c-index, it is possible to test whether the area significantly exceeds 0.5, which would demonstrate that the index test performs better than chance. A worthless test has an area of 0.5. Furthermore, using the SE, also a 95% CI can be derived for the ROC area or c-index. For the D-dimer test example, the area under the ROC curve was 0.86 (SE, 0.011), with a 95% CI of 0.84 to 0.88.

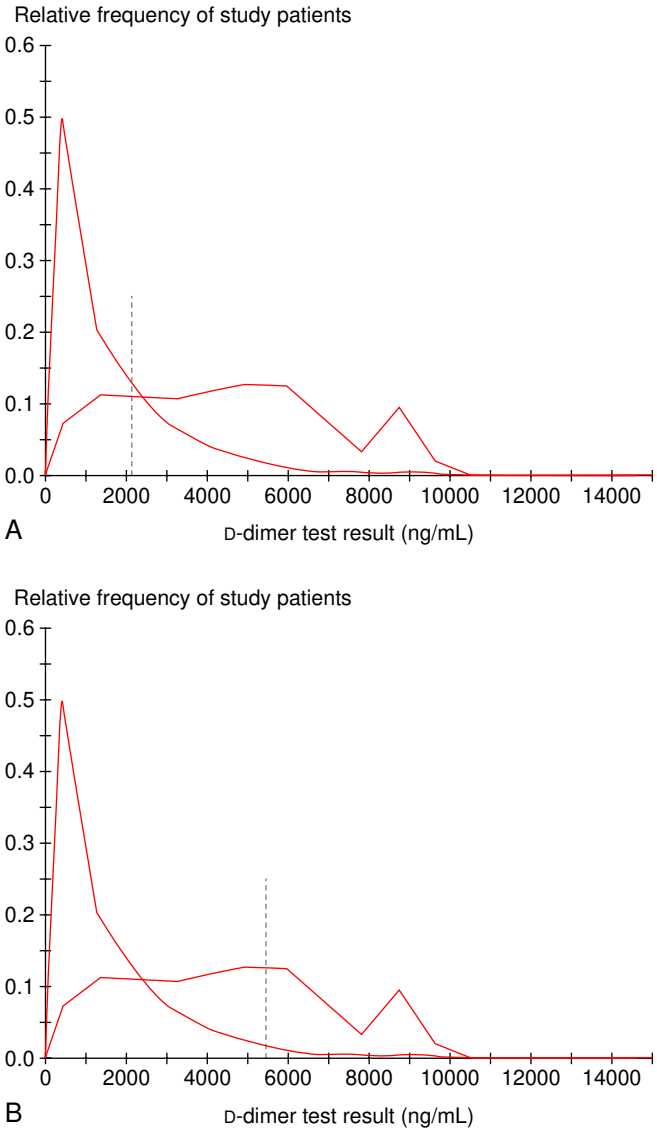


Fig. 2.27 Alternative cutoffs to 500 µg/L in the D-dimer example. (A) Cutoff (2133 µg/L) giving maximum value of the sum of the specificity and sensitivity. (B) Cutoff (5435 µg/L) providing a high specificity (0.975). *Light blue line,* non–deep venous thrombosis (DVT); *blue line,* DVT. The *dashed line* indicates the cutoff value. (From Linnet, K., Bossuyt, P. M., Moons, K. G., & Reitsma, J. B. [2012]. Quantifying the accuracy of a diagnostic test or marker. *Clinical Chemistry, 58,* 1292–1301.)

Selection of Cutoff Value in Case of Quantitative Index Tests

The specificity and sensitivity determined for an index test almost always vary inversely over the range of possible cutoffs. The cutoff point that provides the maximum of the sum of the specificity and sensitivity *could* be selected. In the D-dimer example, this cutoff would be close to 2000 µg/L, yielding a specificity of 0.76 and a sensitivity of 0.80 (Fig. 2.27A). However, this method of cutoff selection is commonly not recommended. The selection should rather be based on the intended purpose of the index test. If an index test is applied primarily to rule out the presence of disease (e.g., in the case of the D-dimer assay for exclusion of DVT), the cutoff point

should be at the lower end of the distribution of values of diseased individuals (see Fig. 2.25) (e.g., a cutoff of 500 µg/L). At this cutoff, the sensitivity approaches 1.0. But attaining such a high sensitivity is at the cost of a loss of specificity because of overlap of test values in the diseased and nondiseased individuals. Conversely, when FP results are judged unacceptable, the cutoff should be toward the upper limit of the distribution of values for the nondiseased group. For the D-dimer test example, a cutoff value corresponding to the 97.5 percentile of the distribution of values for those not having DVT (5435 µg/L) resulted in a specificity of 0.975, but now the sensitivity was only 0.36 (i.e., nearly the opposite of the situation with a cutoff of 500 µg/L) (see Fig. 2.27B).

Posterior Probabilities (Predictive Values)

Unlike sensitivity and specificity, the **positive predictive value** assesses the probability of having the disease given a positive test result, $P(D \mid Tpos)$, whereas the **negative predictive value** assesses the probability of not having the disease given a negative test result $P(Non-D \mid Tneg)$. The probability of presence of target disease given the index test result is an example of a so-called posterior disease probability, where the prior probability corresponds to the prevalence of the disease in the given situation. The **prevalence** of disease (P[D]) in the study sample is the a priori (pretest) probability of disease.

Given a positive test result (Tpos), the posterior disease probability is estimated as the fraction of TP out of all test result positives:

$$P(D \mid Tpos) = TP/(TP+FP)$$

Analogously for a negative result (Tneg), the probability that the given disease is absent is

$$P(Non-D \mid Tneg) = TN/(TN+FN)$$

Just as with sensitivity and specificity values, these posterior disease probabilities depend on the selected cutoff point for a quantitative test. In case of a dichotomous or dichotomized index test, these posterior probabilities are also called predictive values. They are highly dependent on the disease prevalence.

From the Bayes rule, the following relations exist:

$$P(D \mid Tpos) = [Sensitivity \times P(D)] / [Sensitivity \times P(D) + (1-Specificity)(1-P(D))]$$

$$P(Non-D \mid Tneg) = [Sensitivity \times (1-P(D))] / [Specificity \times (1-P(D)) + P(D) \times (1-Specificity)]$$

Likelihood Ratios and Odds Ratios

From relative frequency distributions for results of the index test in the nondiseased and diseased groups, the so-called diagnostic **likelihood ratio** (LR) of an index test result *(X)* can be calculated as the ratio between the heights of the relative frequency *(f)* distributions at that specific test value. We get:

$$LR(X) = f_D(X) / f_{Non-D}(X)$$

In case the relative frequency of the distribution of diseased individuals is higher than that of the nondiseased individuals, the ratio exceeds 1. This indicates that disease is more likely than nondisease given this particular index test result. More formally, the ratio can be used to calculate posterior disease probabilities given specific values of the index test *(X)* and the disease prevalence (D):

$$P(D \mid X) = P(D) \times LR(X) / [P(D) \times LR(X) + (1-P(D))]$$

or a more simple calculation can be carried out using odds instead of probabilities:

$$Odds(D \mid X) = Odds(D) \times LR(X)$$

based on the relation:

$$Odds = P/(1-P)$$

Odds is an alternative way of expressing probabilities commonly used in betting games in Anglo-Saxon countries. For example, a probability of 0.80, or 80%, corresponds to an odds value of 4 according to the formula above. The higher the odds, the closer a probability is to one. From the equation, the posterior odds are equal to the prior odds multiplied by the diagnostic LR for the result *X*.

For a dichotomous or dichotomized index test, the following relationships apply:

$$LR(pos) = Sensitivity/(1-Specificity)$$

$$LR(neg) = (1-Sensitivity)/Specificity$$

A simple way of achieving the posttest probability of disease from the prevalence (pretest probability of disease) and the diagnostic LR is to use the Fagan nomogram. A recent example is the estimation of the probability of DVT from testing for D-dimer.

Comparison of Diagnostic Accuracy of Two Tests in Isolation

The diagnostic accuracy—that is, the ability to detect or exclude the target disease as determined by the reference method—of a new diagnostic index test is usually compared with another, established, index test. We here focus on the pure performances of the tests without consideration of other tests (i.e., we consider each test in isolation). When comparing the accuracy of two or more diagnostic index tests, a paired design is generally preferable for reasons of both validity and efficiency. In the target disease-suspected patients, the two index tests under comparison and the reference standard are performed on all subjects, again independently and blinded with regard to each other's test results. An example of a paired comparison is displayed in Fig. 2.28. Overall, the index test having the largest area under the ROC curve represents the best test. Preferably, CIs of areas and differences of areas should be provided.

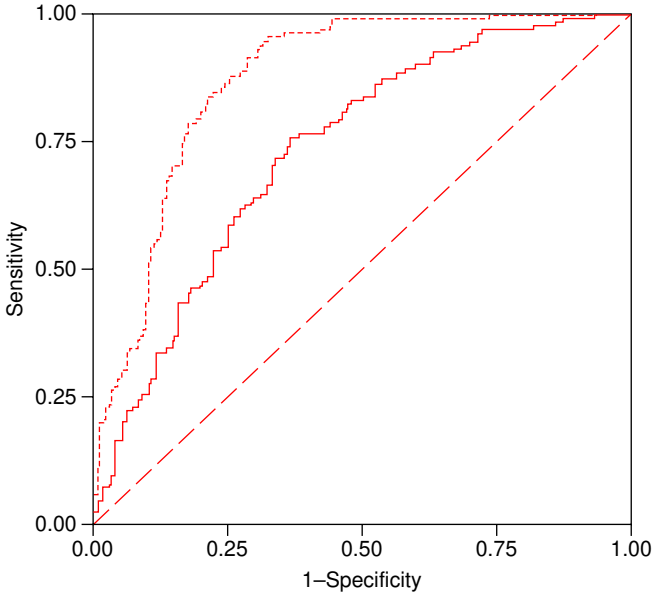


Fig. 2.28 Comparison of the receiver operating characteristic curves of two hypothetical index tests for the same target disease undertaken in the same individuals. The *dotted blue curve* represents a superior diagnostic test, both with regard to sensitivity and specificity over all possible cutoff points. The *dashed diagonal* represents a worthless test, with equal probability of a false-positive (1 – Specificity) and false-negative (1 – Sensitivity) result across all cutoff values (i.e., flipping a coin test). (From Linnet, K., Bossuyt, P. M., Moons, K. G., & Reitsma, J. B. [2012]. Quantifying the accuracy of a diagnostic test or marker. *Clinical Chemistry, 58*, 1292–1301.)

The STARD (Standards for Reporting of Diagnostic Accuracy Studies) initiative (Box 2.2) and the so-called QUADAS-2 (Quality Assessment Tool for Diagnostic Accuracy Studies) aim at improving diagnostic accuracy studies (http://www.quadas.org).

Diagnostic Accuracy of a Test in the Clinical Context

The diagnostic process commonly consists of a series of sequential steps in which much diagnostic information (i.e., diagnostic test results) is acquired. After each step, the physician intuitively judges the probability of the target disease being present. The initial step always consists of patient history and physical signs. If uncertainty about the presence and type of disease remains, subsequent tests are performed, often in another stepwise fashion. These supplementary tests may consist of simple blood or urine tests or be imaging, electrophysiology, or genetic tests or even later in the process more invasive testing such as biopsy, angiography, or arthroscopy. The supplementary information of each subsequent test is implicitly added to the already collected diagnostic information, and the target disease probability is constantly updated. This process continues until the target disease can be included or excluded with sufficient certainty.

Diagnostic test studies should reflect the steps in the diagnostic process so that the added value of such tests in excess of the information that is already present can be assessed. Depending on the situation, studies may reveal that the diagnostic information of any subsequent test is already supplied by the simpler previous test results. When regarded in isolation, such subsequent test or marker may indeed show diagnostic accuracy or value, but when assessed in the overall diagnostic workup, it does not. Such a case can arise because different tests may gauge the same underlying pathologic processes to varying degrees and thus provide related diagnostic information.

Clinical Example: Added Value of D-Dimer Testing in the Diagnosis of Suspected Deep Venous Thrombosis

The same DVT case study described earlier is considered. A total of 2086 patients were suspected of DVT, having at least one of the following symptoms: swelling, redness, or pain in the leg. All patients had a standardized diagnostic workup consisting of index tests from medical history taking, physical examination, and quantitative D-dimer testing. The reference standard was repeated compression ultrasonography, according to current clinical practice, carried out in all patients independent of the results of the index tests and blinded with regard to all preceding collected index test results. In total, 416 of the 2068 included patients (20%) had DVT confirmed by ultrasonography. We focus on estimating the added value of D-dimer testing to the information provided by history taking and physical examination.

A multivariable statistical approach is needed to assess the diagnostic accuracy of combined index test results. Logistic regression models express the probability of DVT (on the logit scale) as a linear function of the included index test results. Note that index test results may be included as binary, categorical, or even continuous results. Table 2.4 (model 1) shows the results from history and physical examination test results that were significantly related to DVT in the multivariable analysis, here defined as a multivariable **odds ratio** significantly ($P < .05$) different from 1 (no association).

To quantify whether the quantitative D-dimer assay result has added diagnostic value beyond the history and physical examination results combined, the basic model 1 was simply extended by including the index test D-dimer value, resulting in model 2 (see Table 2.4). After the inclusion of the D-dimer assay result, the regression coefficients of most history and physical tests in model 2 are found to be different from those in model 1: They now express the contribution of the corresponding test results, given a specific D-dimer result. This change reveals that the history and physical and the D-dimer results are indeed correlated and partly provide the same diagnostic information regarding whether DVT is present or not. The trend of lower regression coefficients of most findings can be interpreted as follows: A portion of the information supplied by the history and physical items is now replaced by the D-dimer assay result.

BOX 2.2 STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies

Identification as a study of diagnostic accuracy using at least one measure of accuracy (e.g., sensitivity, specificity, predictive values, AUC)

Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)

Scientific and clinical background, including the intended use and clinical role of the index test
Study objectives and hypotheses

Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)

Eligibility criteria
On what basis potentially eligible participants were identified (e.g., symptoms, results from previous tests, inclusion in registry)
Where and when potentially eligible participants were identified (setting, location, and dates)
Whether participants formed a consecutive, random, or convenience series

Index test, in sufficient detail to allow replication
Reference standard, in sufficient detail to allow replication
Rationale for choosing the reference standard (if alternatives exist)
Definition of and rationale for test positivity cutoffs or result categories of the index test, distinguishing prespecified from exploratory
Definition of and rationale for test positivity cutoffs or result categories of the reference standard, distinguishing prespecified from exploratory
Whether clinical information and reference standard results were available to the performers or readers of the index test
Whether clinical information and index test results were available to the assessors of the reference standard

Methods for estimating or comparing measures of diagnostic accuracy
How indeterminate index test or reference standard results were handled
How missing data on the index test and reference standard were handled
Any analyses of variability in diagnostic accuracy, distinguishing prespecified from exploratory
Intended sample size and how it was determined

Flow of participants, using a diagram
Baseline demographic and clinical characteristics of participants
Distribution of severity of disease in those with the target condition
Distribution of alternative diagnoses in those without the target condition
Time interval and any clinical interventions between index test and reference standard

Cross-tabulation of the index test results (or their distribution) by the results of the reference standard
Estimates of diagnostic accuracy and their precision (e.g., 95% CIs)
Any adverse events from performing the index test or the reference standard

Study limitations, including sources of potential bias, statistical uncertainty, and generalizability
Implications for practice, including the intended use and clinical role of the index test

Registration number and name of registry
Where the full study protocol can be accessed
Sources of funding and other support; role of funders

AUC, Area under the curve; *CI,* confidence interval; *STARD,* Standards for Reporting of Diagnostic Accuracy Studies.
From Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., et al. (2015). STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Clinical Chemistry, 61*(12), 1446–1452.

Diagnostic Accuracy of Combinations of Diagnostic Tests: Receiver Operating Characteristic Area

The multivariable diagnostic model, which is based on a combination of diagnostic index tests, as exemplified in models 1 and 2 in Table 2.4, can be considered as a single (overall or combined) quantitative index test, consisting of a composite of individual index tests. The test result of this "combined index test model" for each study patient is simply the calculated posterior probability of DVT presence given the observed pattern of the individual index test results in that patient. (See the footnote to Table 2.4 on how to calculate this probability of disease presence.)

As for single continuous index tests described earlier, also for "test combinations" combined into a single multivariable model, the ROC area (*c*-statistic) can be calculated to indicate the ability of this "test combination" to discriminate between

TABLE 2.4 Basic and Extended Multivariable Diagnostic Model to Discriminate Between Deep Venous Thrombosis Presence Versus Absence[a]

	MODEL 1 (BASIC MODEL)			MODEL 2 (BASIC MODEL + D-DIMER)		
	Regression Coefficient (SE)	OR (95% CI)	*P* value	Regression Coefficient (SE)	OR (95% CI)	*P* value
(Intercept)	−3.70 (0.26)	—	<.01	−4.94 (0.32)	—	<.01
Presence of malignancy	0.62 (0.22)	1.9 (1.2–2.9)	<.01	0.22 (0.26)	1.2 (0.7–2.1)	0.41
Recent surgery	0.44 (0.16)	1.6 (1.1–2.1)	<.01	0.003 (0.19)	1.0 (0.7–1.5)	0.99
Absence of leg trauma	0.75 (0.18)	2.1 (1.5–3.0)	<.01	0.67 (0.20)	2.0 (1.3–2.9)	<.01
Vein distension	0.48 (0.13)	1.6 (1.1—2.1)	<.01	0.25 (0.16)	1.3 (0.9–1.8)	0.12
Pain on walking	0.41 (0.15)	1.5 (1.1–2.0)	<.01	0.46 (0.18)	1.6 (1.1–2.3)	0.01
Swelling whole leg	0.36 (0.12)	1.4 (1.1–1.8)	<.01	0.47 (0.14)	1.6 (1.2–2.1)	<.01
Difference in calf circumference (per cm)	0.36 (0.04)	1.4 (1.3–1.5)	<.01	0.29 (0.04)	1.3 (1.2–1.4)	<.01
D-Dimer (per 500 ng/mL)	NA	NA	NA	0.29 (0.02)	1.3 (1.3–1.4)	<.01

[a]Exp (regression coefficient) is the odds ratio (OR) of a diagnostic test result. For example, an odds ratio of 2 for absence of leg trauma (model 2) means that a suspected patient without a recent leg trauma has a two times higher chance of having deep venous thrombosis (DVT) than a patient with a recent leg trauma (because in the latter the leg trauma would more likely be the cause of the presenting symptoms and signs). Similarly, an odds ratio of 1.3 for calf difference in cm (model 2) means that for every centimeter increase in calf circumference difference, a patient has a 1.3 times (or 30%) higher chance of having DVT.

A diagnostic model can be considered as a single overall or combined test consisting of different test results, with the probability of DVT presence as its test result. For example, for a male subject without malignancy, recent surgery, or leg trauma but with vein distension and a painful not swollen leg when walking with a calf difference of 6 cm the formula is (model 1):

$Z = -3.70 + 0.62*0 + 0.44*0 + 0.75*0 + 0.48*1 + 0.41*1 + 0.36*0 + 0.36*6 = -0.65$

The probability for this patient of the presence of DVT based on the basic model then is $\exp(-0.65)/(1 + \exp[-0.65]) = 34\%$.

CI, Confidence interval; *NA,* not applicable; *SE,* standard error.

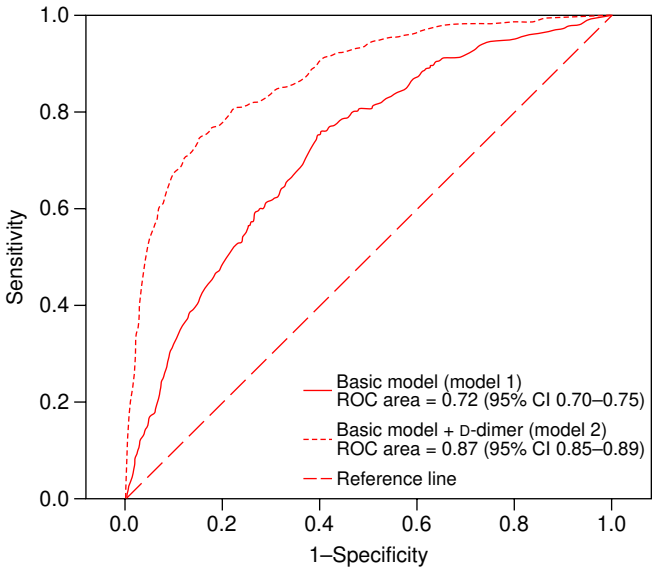


Fig. 2.29 Receiver operating characteristic *(ROC)* curves for the combination of history and physical examination tests before and after addition of the D-dimer assay result. (From Moons, K. G., de Groot, J. A., Linnet, K., Reitsma, J. B., & Bossuyt, P. M. [2012]. Quantifying the added value of a diagnostic test or marker. *Clinical Chemistry, 58,* 1408–1417.)

the presence versus absence of the target disease (here DVT). Fig. 2.29 shows the ROC curves and areas for models 1 and 2. Adding the quantitative D-dimer assay to model 1 mediated an increase in the ROC area from 0.72 to 0.87, a considerable and statistically significant gain ($P < .01$).

Impact of Diagnostic Tests

When thinking about approaches to evaluate the impact of diagnostic tests on medical decision making, patient outcomes, and healthcare at large, it is useful to describe the pathways through which benefits (and risks) of using the test are likely to occur. This so-called *working pathway* provides a framework (Fig. 2.30) to explain how a given test leads to benefits or risks for patients' health or healthcare. Such working pathways include:

1. The anticipated technical or analytical capabilities of the test
2. The unintended and intended results and effects of the test (e.g., benefits of diagnosis and treatment) when applied in the targeted context
3. Those individuals in whom these effects are likely to occur (e.g., in the targeted patients or in the care providers)
4. The anticipated mechanisms through which these potential effects will occur
5. Existing care in the targeted context and individuals
6. The expected time frame in which potential risks and benefits might occur

A clear description of the working pathway of a new test can determine the current benefits (and risks) of prevailing care in the intended medical context. It also helps determine what added value or benefits the new test must provide to improve existing care and what evidence is necessary to quantify whether these (added) benefits are indeed achieved at what risks or costs.

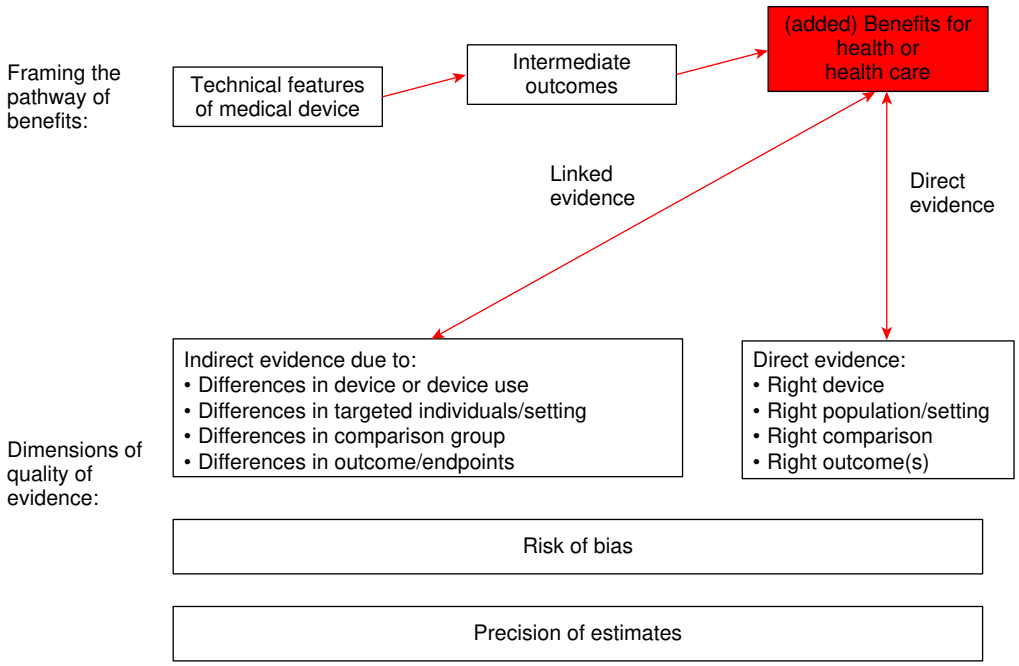


Fig. 2.30 Relationship between the pathway through which devices may lead to benefits or added benefits for health or health care and the three dimensions of quality for evidence (indirectness of evidence, risk of bias, precision of estimates). (From KNAW. [2014]. *Evaluation of new technology in health care. In need of guidance for relevant evidence*. Amsterdam: KNAW.)

POINTS TO REMEMBER

The diagnostic accuracy of a test indicates the frequency and type of errors that a test will produce when differentiating between patients with and without the target disease.

The cohort design based on patients suspected of the diseases targeted by the index test is generally preferable for evaluating diagnostic accuracy.

It is not meaningful to regard estimates of diagnostic performance as properties of the test itself but rather to interpret them as depending on the setting in which the index test was applied and dependent on other tests that are commonly used in that setting.

Focus has been on approaches and measures for quantification of the diagnostic accuracy of combinations of index tests and of the added value of a new diagnostic test beyond existing diagnostic tests.

REVIEW QUESTIONS

1. Which statement applies to a sample of values drawn from a Gaussian distribution?
   - **a.** The central location is best described by the median.
   - **b.** The dispersion is best described by the interquartile range.
   - **c.** The distribution of the values is likely to be asymmetric.
   - **d.** The *t* distribution is useful for estimation of the 95% CI for the mean value.
2. The analytical specificity of an assay is:
   - **a.** the ability of an assay procedure to determine the concentration of a target analyte in the presence of interfering substances in the sample matrix.
   - **b.** the detection limit of a method.
   - **c.** the ability of an analytical method to assess small variations in the concentration of analyte.
   - **d.** the analyte concentration range over which measurements are within the declared tolerances for imprecision and bias of the method.
3. Two analytical methods are to be compared by analysis in parallel of a suitable number of patient samples. Which of the following is correct?
   - **a.** Ordinary least-squares regression analysis is the most appropriate data analysis approach.
   - **b.** It is generally recommended that the manufacturer use 40 samples for comparison and the user laboratory 100 samples.
   - **c.** A calibration difference is most typically disclosed by an intercept estimate significantly different from zero obtained by regression analysis.
   - **d.** In case of constant CV%s, the Bland-Altman difference plot shows an increasing scatter of the measured differences at increasing measurement values.

4. In a regression analysis comparing results of two methods, the *y*-intercept is calculated to be 2.0 and the slope is 3. This indicates a(n):
   a. calibration error.
   b. uncertainty.
   c. systematic difference.
   d. interference in one method.
5. Which one of the following, when stated as an interval around a reported laboratory result, will specify the location of the true value with a given probability?
   a. Traceability
   b. Coefficient of variation
   c. Trueness
   d. Uncertainty
6. The traceability chain extends downwards from the reference measurement procedure to the routine analytical method. Which of the following is correct?
   a. A reference measurement procedure is sensitive to matrix effects.
   b. The standard uncertainty indicates a 95% uncertainty interval.
   c. Harmonization of laboratory measurements do not presuppose traceability to a reference measurement procedure.
   d. The reference measurement procedure is always more precise than the routine analytical method.
7. The diagnostic accuracy of a test is assessed on a number of subjects suspected of having a given target disease. Which of the following is correct?
   a. The diagnostic accuracy is characteristic for the test and is not influenced by the actual setting in which it is evaluated.
   b. The ROC area provides a measure of the diagnostic accuracy, which is not dependent on a selected cut-off value.
   c. When the cut-off value of a quantitative test is increased, the specificity declines and the sensitivity increases.
   d. In order to rule out the presence of disease, it is important that the specificity is high.
8. The probability of the presence of a specific disease divided by the probability of its absence is the:
   a. likelihood ratio.
   b. odds ratio.
   c. prevalence.
   d. predictive value.
9. When a receiver operating characteristic curve is plotted, the *x*-axis represents the:
   a. false-positive rate.
   b. true-positive rate.
   c. false-negative rate.
   d. true-negative rate.
10. A new test is added to an existing set of diagnostic procedures. Which of the following is correct?
   a. The diagnostic accuracy of the new test is the most important point to consider.
   b. A multivariate data treatment based on logistic regression analysis presupposes quantitative test results.
   c. It is unlikely that results from several tests are correlated.
   d. The difference between the ROC curve area after addition of the new test and the area of the ROC curve of the original diagnostic procedure expresses the added value of the new test.

SUGGESTED READINGS

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *i*, 307–310.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., et al. (2015). An updated list of essential items for reporting diagnostic accuracy studies. *Clinical Chemistry*, *61*, 1446–1452.

Bossuyt, P. M. M., Reitsma, J. B., Linnet, K., et al. (2012). Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clinical Chemistry*, *58*, 1636–1643.

Dybkær, R. (1997). Vocabulary for use in measurement procedures and description of reference materials in laboratory medicine. *European Journal of Clinical Chemistry and Clinical Biochemistry*, *35*, 141–173.

Hendriksen, J. M. T., Geersing, G. J., van Voorthuizen, S. C., et al. (2015). The cost–effectiveness of point-of-care D-dimer tests compared with a laboratory test to rule out deep venous thrombosis in primary care. *Expert Review of Molecular Diagnostics*, *15*, 125–136.

Horvath, A. R., Lord, S. J., StJohn, A., et al. (2014). From biomarkers to medical tests: the changing landscape of test evaluation. *Clinical Chemistry Acta*, *2014*(427), 49–57.

Krouwer, J. S. (2002). Setting performance goals and evaluating total analytical error for diagnostic assays. *Clinical Chemistry*, *48*, 919–927.

Linnet, K., Bossuyt, P. M., Moons, K. G., et al. (2012). Quantifying the accuracy of a diagnostic test or marker. *Clinical Chemistry*, *58*, 1292–1301.

Linnet, K. (1990). Estimation of the linear relationship between the measurements of two methods with proportional errors. *Statistics in Medicine*, *9*, 1463–1473.

Linnet, K. (1993). Evaluation of regression procedures for methods comparison studies. *Clinical Chemistry*, *39*, 424–432.

Linnet, K. (1999). Limitations of the paired t-test for evaluation of method comparison data. *Clinical Chemistry*, *45*, 314–315.

Linnet, K. (2000). Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. *Clinical Chemistry*, *46*, 867–869.

Moons, K. G., de Groot, J. A., Linnet, K., et al. (2012). Quantifying the added value of a diagnostic test or marker. *Clinical Chemistry*, *58*, 1408–1417.

Moons, K. G. (2010). Criteria for scientific evaluation of novel markers: a perspective. *Clinical Chemistry*, *56*, 537–541.

Obuchowski, N. A., Lieber, M. L., & Wians, F. H., Jr. (2004). ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clinical Chemistry*, *50*, 1118–1125.

Oudega, R., Moons, K. G., & Hoes, A. W. (2005). Ruling out deep venous thrombosis in primary care. A simple diagnostic algorithm including D-dimer testing. *Thrombosis and Haemostasis*, *94*, 200–205.

Passing, H., & Bablok, W. (1984). Comparison of several regression procedures for method comparison studies and determination of sample sizes. *Journal of Clinical Chemistry and Clinical Biochemistry*, *22*, 431–445.

Pencina, M. J., D'Agostino, R. B., & Vasan, R. S. (2010). Statistical methods for assessment of added usefulness of new biomarkers. *Clinical Chemistry and Laboratory Medicine*, *48*, 1703–1711.

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed.). Ames, Iowa: Iowa State University Press, 75. 121, 140–142, 170–4, 177, 237–8, 279.

Vesper, H. W., & Thienpont, L. M. (2009). Traceability in laboratory medicine. *Clinical Chemistry*, *55*, 1067–1075.