EP21-A
Vol. 23 No. 20
Replace EP21-P
Vol. 22 No. 2

Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline

This document provides manufacturers and end users with a means to estimate total analytical error for an assay. A data collection protocol and an analysis method which can be used to judge the clinical acceptability of new methods using patient specimens are included. These tools can also monitor an assay's total analytical error by using quality control samples.

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

NCCLS

NCCLS...

Serving the World's Medical Science Community Through Voluntary Consensus

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

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

PUBLICATIONS

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

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

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

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

CONSENSUS PROCESS

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

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

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

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

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

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

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

COMMENTS

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

VOLUNTEER PARTICIPATION

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

EP21-A
ISBN 1-56238-502-X
Volume 23 Number 20 ISSN 0273-3099

Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline

Jan S. Krouwer, Ph.D., Chairholder
J. Rex Astles, Ph.D., FACB
W. Gregory Cooper
Steven I. Gutman, M.D., MBA
David D. Koch, Ph.D.
Jacob B. Levine, M.B.A.
Catherine N. Otto, Ph.D.

Abstract

NCCLS document EP21-A—*Estimation of Total Analytical Error for Clinical Laboratory Methods*; *Approved Guideline* provides manufacturers and end users with a means to estimate total analytical error for an assay. The procedure includes a data collection protocol and analysis method, which is largely graphical. The result is compared to a total analytical error goal.

The procedure described herein relies on a method comparison data collection procedure such as that described in NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples*. Two calculation methods are used: parametric and nonparametric. NCCLS document EP21 can be used to judge the clinical laboratory acceptability of new methods by using patient specimens, or to monitor an assay's total analytical error by using quality control samples.

NCCLS. *Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline*. NCCLS document EP21-A (ISBN 1-56238-502-X). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2003.

Suggested Citation

(NCCLS. *Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline.* NCCLS document EP21-A [ISBN 1-56238-502-X]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2003.)

Proposed Guideline
January 2002

Approved Guideline
April 2003

ISBN 1-56238-502-X
ISSN 0273-3099

Committee Membership

Area Committee on Evaluation Protocols

Jan S. Krouwer, Ph.D.
Chairholder
Krouwer Consulting
Sherborn, Massachusetts

Daniel W. Tholen, M.S.
Vice-Chairholder
Statistical Services
Traverse City, Michigan

Subcommittee on Total Error

Jan S. Krouwer, Ph.D.
Chairholder
Krouwer Consulting
Sherborn, Massachusetts

J. Rex Astles, Ph.D., FACB
Centers for Disease Control and Prevention
Atlanta, Georgia

W. Gregory Cooper
Bio-Rad Laboratories
Irvine, California

Steven I. Gutman, M.D., MBA
FDA Ctr. for Devices/Rad. Health
Rockville, Maryland

David D. Koch, Ph.D.
Merck Research Laboratories, Merck & Co.
West Point, Pennsylvania

Jacob B. Levine, M.B.A.
Bayer Corporation
Tarrytown, New York

Catherine N. Otto, Ph.D.
University of North Carolina – Chapel Hill
Chapel Hill, North Carolina

Advisors

W. Gregory Miller, Ph.D.
Virginia Commonwealth University
Richmond, Virginia

Donald R. Parker, Ph.D.
Bayer Corporation
Elkhart, Indiana

Max Robinowitz, M.D.
FDA Ctr. for Devices/Rad. Health
Rockville, Maryland

Leticia J. San Diego, Ph.D.
Health Care Management Development Ctr.
Clinton Township, Michigan

Manoj Sharma
LifeScan
Milpitas, California

Staff

Lois M. Schmidt, D.A.
Staff Liaison
NCCLS
Wayne, Pennsylvania

Patrice E. Polgar
Editor
NCCLS
Wayne, Pennsylvania

Donna M. Wilhelm
Assistant Editor
NCCLS
Wayne, Pennsylvania

Contents

Abstract i

Committee Membership iii

Foreword vii

1 Scope 1

2 Introduction 1

3 Definitions 1

4 General Principles 3

4.1 Cause and Effect Diagram of Medical Errors 3
4.2 The Importance of Total Analytical Error 5
4.3 Obtaining Total Analytical Error Goals 5
4.4 Different Uses of Total Analytical Error 5

5 Protocols 6

5.1 Using NCCLS Document EP9—*Method Comparison and Bias Estimation Using Patient Samples* as the Basis for Collecting the Data 6
5.2 Using Quality Control Samples 6

6 Data Analysis 7

6.1 Graphical Analysis 7
6.2 Total Analytical Error Estimation: Introduction 8
6.3 Total Analytical Error Estimation: Outliers 9
6.4 Total Analytical Error Estimation: Nonparametric Analysis 9
6.5 Parametric Analysis 10
6.6 A Review of Modeling Approaches 10
6.7 An Outline of a Simulation Method to Estimate Total Analytical Error 11
6.8 Comparison of the Total Analytical Error Evaluation Protocol With Other Evaluation Methods 11

7 Presentation of Results with Examples 11

7.1 Result Report Format 11
7.2 Result Report Interpretation 12
7.3 Example 1. LDL Cholesterol 12
7.4 Example 2. Sodium 18

References 24

Appendix A. Factors (k) to Calculate Normal Distribution Tolerance Intervals 25

Appendix B. Number of Extreme Observations ν to Be Removed from the Ends of a Sample Size *n* to Obtain a Two-Sided Distribution-Free Tolerance Interval or to Obtain a One-Sided Distribution-Free Tolerance Bound That Contains at Least 100*p*% of the Sample Population with 100(1-α)% Confidence 26

Appendix C. A Mathematical Representation of Total Analytical Error 27

Contents (Continued)

Summary of Consensus Comments and Committee Responses 28

Summary of Delegate Comments and Subcommittee Responses 33

The Quality System Approach 35

Related NCCLS Publications 35

Foreword

NCCLS document EP21—*Estimation of Total Analytical Error for Clinical Laboratory Methods* provides manufacturers and end users a means to estimate total analytical error for an assay. The procedure includes a data collection protocol and analysis method, which is largely graphical. The result is compared to a total analytical error goal.

Total analytical error (or measurement error) refers to assay errors from all sources arising from the data collection experiment. If the data collection protocol is representative of routine assay use, then estimation of total analytical error will provide a snapshot of the distribution of differences between a candidate assay and its comparison method. If the comparison method is a reference method and the reference method imprecision is minimized through replication, then except for the small imprecision left in the reference method, these differences estimate differences between the candidate method's result and the true assay concentration. When the candidate method is a comparison method, these differences cannot all be attributed to the candidate method, yet they are nevertheless important to estimate, since they will be observed by clinicians.

Whereas total analytical error is of paramount importance in judging the acceptability of a candidate method, it is surprising how infrequently it has been used. Alternatives to total analytical error have been the calculation of total error sources and their combinations ("estimation of measurement uncertainty"). This is a complex procedure which requires correctly specifying a model that describes how the error sources should be combined. Another frequent alternative has been estimation of some of the total analytical error sources without any attempt to combine them.

The procedure described in this guideline relies on a method comparison data collection procedure such as that described in NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples*. Two calculation methods are used: parametric and nonparametric. Whereas the parametric method is more efficient statistically, it requires normally distributed differences. This can be achieved by transformation; however, the committee decided that the transformation process and its associated tests would be too complicated for many users. The nonparametric method has no assumptions about the distribution. Two plots are produced to help users assess the distribution of differences.

This guideline can be used to judge the clinical laboratory acceptability of new methods by using patient specimens, or to monitor an assay's total analytical error by using quality control samples.

A Note on Terminology

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

In the context of this document, it is necessary to point out that the term *"Total analytical error*" is used differently in the U.S. than in other countries, notably those in Europe. *"Total analytical error"* is an accepted U.S. term and is used in EP21-A the way *"Total analytical error interval"* would be conceptualized in Europe, due to its description of a distribution of errors. At this time, the subcommittee has chosen *not* to replace the term due to U.S. nonfamiliarity and for the sake of the practicability of the guideline.

There are several cases in this guideline, however, where *"Total analytical error"* corresponds to the ISO term *"Error of measurement,"* i.e., the "result of a measurement minus a true value of the measurand." (VIM93-3.10). Both expressions contain both random and systematic effects. Users of EP21-A should understand that the fundamental meanings of these terms are similar, and where appropriate, the ISO term has been inserted parenthetically throughout the text. All terms are also defined along with explanatory notes in the guideline's Definitions section.

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

Key Words

Error, error of measurement, measurement error, total analytical error, total analytical error interval

Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline

1 Scope

This document presents protocols and procedures to estimate and report analytical error for clinical laboratory assays. These protocols and procedures are applicable to all quantitative analytical clinical laboratory methods. The target audience includes both manufacturers and end users in laboratories. The prime procedural difference between these two groups is the amount of data collected.

This guideline will be useful to users in settings with only one instrument as well as to those with multiple laboratories and multiple instruments.

2 Introduction

Total analytical error as a concept has been around for many years.[1,2] However, the use of total analytical error as an evaluation method for diagnostic assays is more recent.[3,4] As Westgard pointed out in his 1974 article, physicians, prime consumers of diagnostic test data, think in terms of total analytical error, not random or systematic error. Yet for many years, the estimation of random and systematic error was often performed without an attempt to combine them to estimate total analytical error, the real metric of interest.

This document addresses this deficiency by providing a method to directly estimate total analytical error, both for manufacturers and end users in laboratories. To facilitate understanding of the interrelationship of all error sources, a hierarchical, graphical chart known as a "cause and effect diagram" is used. It is recommended that for most cases, if one has knowledge of total analytical error and outliers, then one has sufficient information to judge the acceptability of a diagnostic assay.

A data collection protocol is presented for patient samples that leverages existing NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples*. Quality control data can be used to monitor total analytical error, after estimation of total analytical error with patient specimens.

Two calculation methods are presented: a parametric method which requires certain assumptions, and a nonparametric method which requires fewer assumptions. Both methods are supported by graphical analyses that can be performed independently of any calculations. Estimation requires total analytical error specifications. This document assumes these specifications exist and does not address their generation.

Finally, this guideline briefly reviews modeling methods and outlines a simulation approach. The simulation method is more suitable for manufacturers due to its complexity.

Examples illustrate these concepts with actual data.

3 Definitions

Accuracy - Closeness of the agreement between the result of a measurement and a true value of the measurand (*VIM93-3.5*.)

Assay - A quantitative determination or measurement of the amount, activity, or potency of a constituent or characteristic.

Bias - The difference between the expectation of the test results and an accepted reference value (*ISO 3534-1; 1993*)[5]; NOTE**:** In general, the deviation/difference is based on replicate measurement using an accepted (definitive, reference, or designated comparison) method and the method being tested, and expressed in the units of the measurement or as a percentage.

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

Coefficient of variation, CV - For a non-negative characteristic, the ratio of the standard deviation to the average; **NOTE:** It is often multiplied by 100 and expressed as a percentage (*ISO 3534-1; 1993*)[5]

Concentration - A measure of the amount of dissolved substance per unit of volume.

Control//Control material - A device, solution, or lyophilized preparation intended for use in the quality control process; **NOTES:** a) The expected reaction or concentration of analytes of interest are known within limits ascertained during preparation and confirmed in use; b) Control materials are generally not used for calibration in the same process in which they are used as controls.

Cumulative distribution - For any probability distribution, the cumulative distribution represents the set of each ordered value of the variable with its corresponding percentile; **NOTES:** a) An empirical cumulative distribution assumes that the underlying distribution is unknown; b) For example, if there were five sodium values of 140, 142, 139, 150, and 145, the empirical cumulative distribution is: {value, percentile;}: {139, 17%; 140, 33%; 142, 50%; 145, 67%; 150, 83%;}. In this example, the percentile 67% means that there is a probability of 67% for this distribution that values will be below 145.

Error - 1) Deviation from truth or from an accepted, expected true or reference value; **2) Random error** - result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions [*VIM; 3.13*]; **3) Systematic error -** mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand [*VIM; 3.14*]; **NOTES:** a) Systematic error is equal to error minus random error; b) Like the true value, systematic error and its causes cannot be completely known.

Error of measurement//Measurement error - Result of a measurement minus a true value of the measurand (*VIM 93-3.10*).

Inaccuracy - The numerical difference between a value and the true value.

Measurement uncertainty//Uncertainty of measurement - A parameter, associated with the result of measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand [*VIM93-3.9*]; **NOTE:** The parameter may be, for example, a standard deviation (or given multiple of it), or the half width of an interval having a stated level of confidence.

Measuring system - A complete set of measuring instruments and other equipment assembled to carry out specified measurements.

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

{Result//Measurement result}//Result of a measurement - The value attributed to a measurand obtained by measurement.

Sample - One or more parts taken from a system and intended to provide information on the system, often to serve as a basis for decisions on the system or its production (*ISO DIS 15190*).

Specimen - The discrete portion of a body fluid or tissue taken for examination, study, or analysis of one or more quantities or characteristics to determine the character of the whole.

Target value - ***In Quantitative Testing,*** either the mean of all participant responses after removal of outliers (those responses greater than three standard deviations from the original mean) or the mean established by definitive or reference methods acceptable for use in the National Reference System for the Clinical Laboratory (NRSCL) by the National Committee for Clinical Laboratory Standards (NCCLS); NOTE**:** In instances where definitive or reference methods are not available or a specific method's results demonstrate bias that is not observed with actual patient specimens as determined by a defensible scientific protocol, a comparative method or a method group ("peer" group) may be used. If the method group is less than ten participants, "target value" means the overall mean after outlier removal (as defined above) unless acceptable scientific reasons are available to indicate that such an evaluation is not appropriate.

Test - ***In the Clinical Laboratory***, a qualitative, semiqualitative, quantitative, or semiquantitative procedure for detecting the presence of, or measuring the quantity of an analyte.

Tolerance interval - A range that contains a specified proportion of a sampled population with a specified confidence.

Total analytical error - In the context of this guideline, "total analytical error" is used to describe the following concepts: **1)** the interval that contains a specified proportion (usually 90, 95, or 99 %) of the distribution of differences in concentration between the test and reference method; **NOTE a):** For example, 97.2% of the differences between the test and reference method fell within the limits of ±4 mmol/L, hence the 95% total analytical error goal was met; and **2)** the "result of a measurement minus a true value of the measurand," which is the VIM (93-3.10) definition of the term *"Error of measurement";* **NOTE: b)** Both "total analytical error" and "error of measurement" contain random and systematic effects. **NOTE**: **c)** To facilitate understanding, when "total analytical error" is used this way, "measurement error" has been inserted parenthetically throughout the text.

Validation - Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled (*ISO 9000:2000*).

4 General Principles

4.1 Cause and Effect Diagram of Medical Errors

A cause and effect diagram, similar to a fault tree or hazard analysis, is a hierarchical chart whereby a top-level (fault) event is caused by the events connected below it. These "child" events are also connected to events below them, which are in turn, their causes.

For this document, the top-level event is morbidity/mortality, which is the most undesirable event in a medical intervention. Naturally, there are many causes for this top-level event; these are not explored in our cause and effect diagram, since they are beyond the scope of this document. The practical starting point of the cause and effect diagram is the laboratory error block.

Pre- and postanalytical error and outliers are sources of laboratory error. They are not discussed further, since this document focuses on total analytical error.

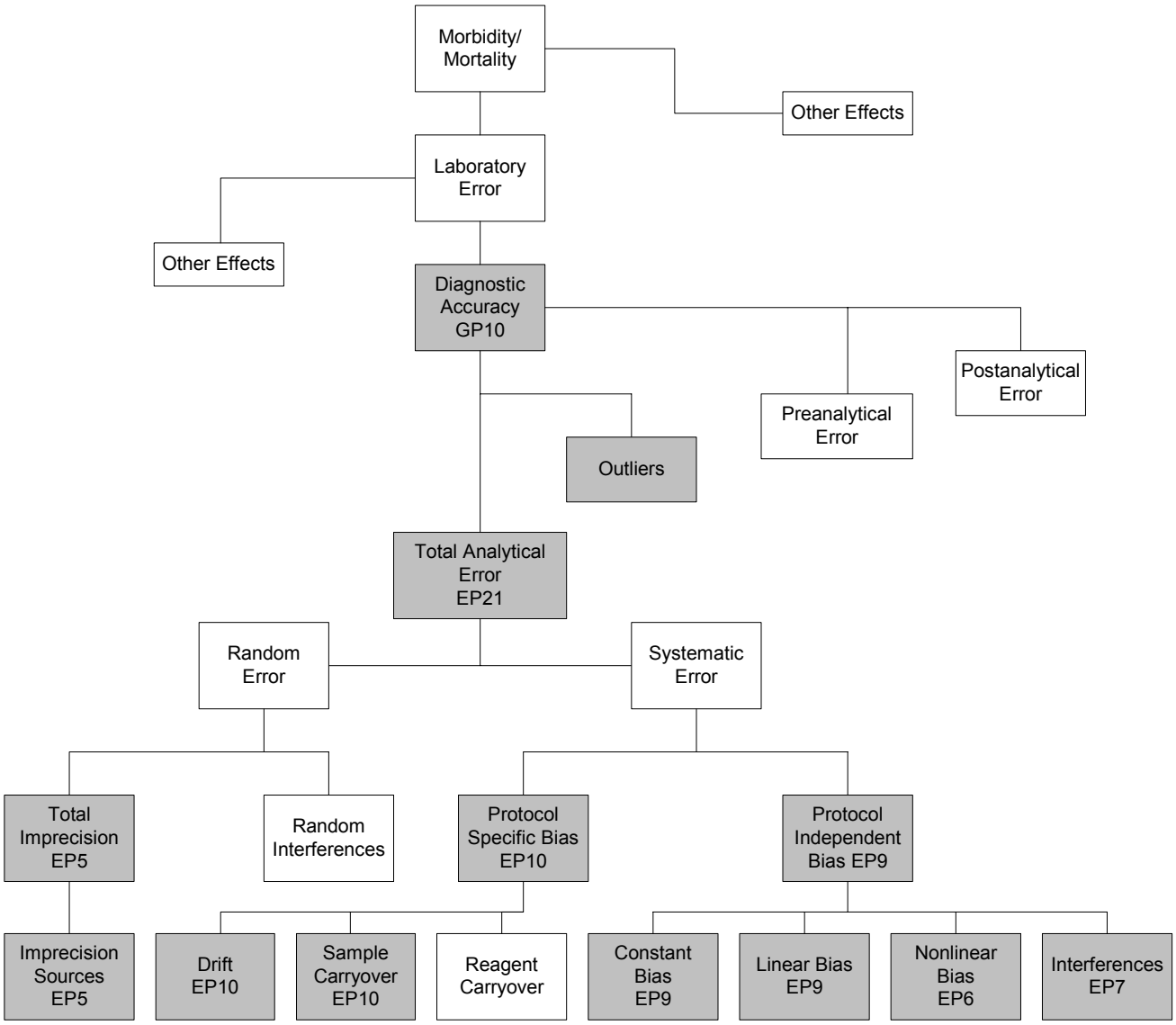

Figure 1. Cause and Effect Diagram of Total Analytical Error (NOTE: "EP" and "GP" numbers refer to NCCLS Evaluation Protocols and General Laboratory Practices documents that recommend a method to estimate that error source.) See Section 6.3 for a discussion of outliers.

Diagnostic accuracy is also not discussed. This error source is beyond the scope of this document and covered in NCCLS document GP10—*Assessment of the Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plots*. Finally, the total analytical error box is intended to cover any deviation of a result from its expected value, for those causes due to some failure in the analytical system (instrument/reagents/software). Outliers and total analytical error (measurement error) must be considered at the same time. Both of these error sources are necessary because of the possible distribution of errors. For example, one could have no outliers but still have unacceptable total analytical error because too many results are outside of the total analytical error specifications, or one could pass total analytical error but still have outliers. Finally, the boxes below total analytical error show how the traditionally estimated error sources are related to total analytical error (measurement error) and thus contribute to the total analytical error interval.

NOTE: In Figure 1, "EP" or "GP" followed by a number refers to an NCCLS document that provides a method to estimate that error source.

The purpose of this document is to provide a method to directly estimate total analytical error, without having to estimate and combine all of the error sources below total analytical error. This obviates a rather complicated and time-consuming process to provide the desired metric of total analytical error more economically.

4.2 The Importance of Total Analytical Error

To understand the importance of total analytical error, one must understand how diagnostic assays are used. Test results are part of a clinician's decision to treat a patient. In some cases, the test results play a major role. For example, the decision to perform a prostate biopsy is largely based on a prostate-specific antigen (PSA) result. If the reported result is in error, the reason for the error is of no practical importance to the clinician. He or she has made a treatment decision, based on an incorrect assay result.

No assay is perfect with respect to error, since assays rely on a combination of complicated instrumentation and sensitive reagents. Yet, if one hopes to reduce medical errors, one would like to specify a simple metric for all error sources and provide a means to measure that source. The value of total analytical error is that it provides a simple measure of assay quality that can be directly equated with medical errors. Thus, total analytical error is a measure of importance for clinicians. Individual sources of imprecision and bias that comprise total analytical error (measurement error) are important for manufacturers to understand, since their improvement is the only means to reduce total analytical error.

4.3 Obtaining Total Analytical Error Goals

Manufacturers provide claims for their assays; yet these claims are often not stated in total analytical error terms. Moreover, many laboratories, while having process control limits, also do not have medically based total analytical error limit. If a laboratory does not have a total analytical error limit, it does not mean that there are no limits (there are) or that not meeting these limits will not cause problems (they can). It simply means that the laboratory does not know what they are.

Laboratories should thus attempt to establish limits, for which there are many strategies suggested in the literature.[6] Unfortunately, most of these strategies describe separate limits for bias and imprecision rather than for total analytical error directly. An International Organization for Standardization (ISO) document summarizes approaches.[7]

In the absence of these limits, one can nevertheless measure the observed total analytical error and perhaps relate the results to the frequency of complaints or medical errors to assess whether the observed results are adequate.

4.4 Different Uses of Total Analytical Error

There are two main uses for the characterization of the differences between two measurement procedures.

In the first, one wants to know the differences that can be expected when one switches from one commercial test method to another. In this case, differences cannot be ascribed to the new method if the current method is not a reference method. Hence the term "difference" is more appropriate than (total analytical) "error."

In the second case, one wants to know the difference between a candidate method and a reference method. Here, it is assumed that the reference method has little or no bias so one can attribute the difference to error in the new method, with one caution. Since the calculations described in this document include imprecision in the comparison (reference) method as part of the difference, the reference method must be replicated and the average value used in the calculation of differences. This will minimize the contribution of random error from the reference method in the estimate of total analytical error. To

determine the number of replicates, one needs to have an estimate of imprecision for the reference method. In general, if the random error (as a CV or standard error) in the reference value (mean of replicates) is less than one third of the random error in the candidate method result, then the additional error is not important.

One should be aware that the relative contribution to the total analytical error estimate of imprecision in the reference assay depends on all error sources, most of which are unknown.

In case 1 (no reference method available), replication is not needed, since we are interested in differences between the two methods regardless of their origin.

5 Protocols

5.1 Using NCCLS Document EP9—*Method Comparison and Bias Estimation Using Patient Samples* as the Basis for Collecting the Data

NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples* is an existing guideline for method comparisons. This is a useful protocol for collecting data for total analytical error estimation. The minimum sample size recommended in NCCLS document EP9 is 40. That is, 40 patient specimens should be run by each assay. Although EP9 requires duplicates to be run for each assay, duplication is not needed for total analytical error estimation, unless one wishes to minimize imprecision in a reference method, whereby one should run duplicates (or higher numbers of replicates) on the reference specimens to minimize imprecision. Because setting up a protocol is often much of the expense of a study, users should consider increasing the number of specimens run beyond the bare minimum. It is important to have representative samples, and this is more likely to occur with larger samples.
For estimation of total analytical error by manufacturers, a recommended sample size is 120. However, manufacturers often analyze many more samples (both in-house and in clinical trials) and are encouraged to use all of this data.

The sample concentrations should span the range of the assay as recommended in NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples*. However, it may be appropriate to have different goals for different concentration ranges. For example, it is well known that glucose has different medical goals at different concentration levels.[8] A separate protocol should be conducted for each concentration range corresponding to each total analytical error goal. Ideally, 120 samples should be used for each concentration range, although this may not be practical for all concentration ranges.

All specimens should be included in the analysis if they are representative of specimens that will be sampled in routine use. Thus, if a sample has icterus and would be routinely analyzed, it should be part of the study. If the sample interferes, this interference will and should appear as part of total error. If it is known that a sample has an interfering substance that exceeds the manufacturers recommendation, and if this knowledge is available for routinely assayed samples, then this sample should be excluded from the analysis. Fresh samples should be used rather than stored samples. However, stored samples may be used as a practical convenience, if the storage conditions have been shown not to affect any aspect of the assay.

5.2 Using Quality Control Samples

As described in other sections of this guideline, total analytical error (measurement error) consists of all possible error sources. This means that evaluating total analytical error *must* be based on analysis of patient specimens. However, after an initial evaluation has been conducted, total analytical error may be monitored by analyzing quality control materials, which are routinely assayed anyway. Users should be aware, however, that total analytical error might be potentially underestimated with quality control samples. This underestimation will depend on the actual error source contributions, for which one has no knowledge.

5.2.1 Additional Information About the Use of Control Materials

Follow the control manufacturer's instructions explicitly on all occasions regarding storage and handling. When using lyophilized controls, be sure to follow manufacturer instructions regarding reconstitution of the control material, particularly when distilled or deionized water and use of a volumetric pipette is recommended.

Use control products that have both normal and clinically relevant concentrations. Depending on the analyte, these abnormal levels may be high or low (or toxic in the case of therapeutic drugs).

Regular calculation of total analytical error estimates with comparison to total analytical error budgets can be valuable to tracking the *long-term* performance of an instrument. Such estimates can alert the operator when one or more of the components of the test system either deteriorate or are changed.

6 Data Analysis

Assumptions: The following assumptions are made in writing the two subsections (e.g., nonparametric analysis and parametric analysis):

- The total analytical error is defined as an interval that contains a specified proportion (usually 90, 95, or 99%) of the distribution of differences in concentration between the test and reference method.[4]

- The data for determining this interval is obtained from an experiment designed using sound statistical principles (i.e., representative of population) and including the sources of variation that truly represent the total system error.

- The differences may be concentration dependent, i.e., the interval may be different at different concentrations. It is assumed that this aspect has been addressed by suitably including different concentrations in the experiment.

6.1 Graphical Analysis

Given that one has a set of differences between a new method and a comparison method, there are two graphs that should be prepared to assess total analytical error: a difference plot and a mountain plot. Each conveys different perspectives about the data.

6.1.1 The Difference Plot

For examples of difference plots refer to Figures 2 and 4.

6.1.1.1 Preparation

The difference plot is simply a plot of the differences between the new and comparison method (y-axis) and either the comparison method or the mean between the new and comparison methods (suggested by Bland and Altman)[3] (x-axis). See also the most current version of NCCLS document EP9—*Method Comparison and Bias Estimation Using Patient Samples.* Bland and Altman offer statistical reasons for choosing the mean of both methods, *when* the comparison method is not a reference method. For assays that have a large range, percent differences might be a better choice, especially if total analytical error is specified as deviation from a stated percent error. Alternatively, one could prepare different plots for different regions of the assay. To help judge acceptability, one can add horizontal specification lines to the y-axis that correspond to total analytical error limits.

6.1.1.2 Interpretation

The width of the data band is a rough indication of the total analytical error (measurement error) relative to the comparative method. One must use caution with respect to interpreting differences if the comparative method is not a reference method, as differences may be due to the comparative method. Also, for either a comparative or reference method, some of the differences will be due to imprecision in the comparative or reference method. One should also look for other trends in the data. Increasing scatter would indicate that random error changes with concentration, leading to higher total analytical error (measurement error.) If the data band were curved, this would indicate that there is a nonlinear component in total analytical error (measurement error).

6.1.2 The Mountain Plot

For examples of mountain plots refer to Figures 3 and 5.

6.1.2.1 Preparation

The mountain plot[8,9] more formally is known as a "folded empirical cumulative distribution." It is prepared starting from the same variable as used in preparing a difference plot, namely differences or percent differences between the new and comparison method or between the new method and the average between the new and comparison method. These differences are then ranked and converted to percentiles, where the percentile = rank x 100 / (N + 1). In this formula, 1 is added to N by convention for empirical, sample percentiles. To get a folded plot, one performs the following transformation for all percentiles over 50: percentile = 100 – percentile. The percentiles (y-axis) are plotted against either the differences or percent differences (x-axis).

To facilitate interpretation, one can add specification lines. For example if the percent total analytical error is set at 95%, one can add a specification line at the 2.5th percentile. One can also add vertical lines corresponding to the total analytical error limits at ± levels of either the difference or percent difference.

6.1.2.2 Interpretation

If the mountain plot curve is contained within the intersection between the horizontal and vertical specification lines, then the total analytical error goal has been met. If a portion of the curve is outside of this intersection, one can read where the curve intersects the horizontal line to see how much beyond the specification the corresponding difference (or percent difference) is.

6.2 Total Analytical Error Estimation: Introduction

Typically, a subcommittee would decide among all estimation procedures and select one. In this case two estimation procedures are provided. The parametric procedure requires normally distributed differences. If the differences are not normally distributed, they must be transformed. The subcommittee felt this procedure was too complicated for many users. The nonparametric procedure requires no assumptions about the distribution of differences. If the differences are normally distributed, both procedures should give similar results.

Given that one has calculated total analytical error estimates, one should realize that these estimates have uncertainty. One can account for this uncertainty by calculating tolerance intervals. Tolerance intervals are wider than the initial total analytical error intervals calculated in Sections 6.4 and 6.5. The extra width that the tolerance intervals impart to the total analytical error intervals is inversely related to the sample size of the experiment (i.e., the larger the sample size, the smaller the interval will be widened). One must be aware, however, that tolerance intervals will only be correct if the sample taken is *representative*. This assumption is difficult to prove, especially for small sample sizes.

6.3 Total Analytical Error Estimation: Outliers

Outliers are observations that appear to be detached from the rest of the distribution of differences. They should be verified to determine whether or not they are mistakes such as mismatched samples (e.g., a difference not from the same patient). There is no statistical basis for removing data before estimating total analytical error. Thus, unless the detached observation(s) are mistakes, they must be kept.

An additional consideration is the relationship between outliers and total analytical error (measurement error). In this document, outliers are considered values beyond predefined limits. Typically, these limits will be wider than total analytical error limits. Note that this definition means that results that are not detached from a distribution might still be considered outliers. To evaluate outliers, one estimates the rate at which they occur. To pass an outlier goal, the observed rate must be below the specified rate. Table 1 shows the possible outcomes of a study to evaluate total analytical error and outliers.

Table 1. Evaluation Results for Total Analytical Error and Outliers

Case	Evaluation of Total Analytical Error	Evaluation of Outliers
1	Pass	Pass
2	Pass	Fail
3	Fail	Pass
4	Fail	Fail

Case 2 implies that while the specified proportion of the distribution of differences is within the total analytical error limits, there are too many results (outliers) that will cause problems.

Case 3 implies that the distribution of differences is not within the total analytical error limits, even though the outlier rate is acceptable.

6.4 Total Analytical Error Estimation: Nonparametric Analysis

The nonparametric analysis does not assume any particular distribution for the population of differences. However, it assumes random sampling and the inclusion of pertinent sources of variation in the experiment. This analysis is more robust, because it makes minimal assumptions.

Let us consider an experiment with n observations on the test and reference method. The difference between the test and the reference method is defined as error. Interest lies in understanding the distribution of this error. We have n observed differences, i.e., the sample size is n. The n differences can be arranged in ascending order. Let these ordered differences be denoted by $x_{(1)}, x_{(2)}, \ldots\ldots, x_{(n)}$. Ranks are computed for each $x_{(n)}$. Each ordered difference is equivalent to a rank. For example, rank 2 = $x_{(2)}$. Ties are given the same rank. Example: ordered results=102, 103, 103, 105, 106. Ranks=1, 2, 2, 4, 5. Percentiles are then computed for each rank according to percentile = rank x 100 / (n +1), where n = the total number of differences and n +1 is used to reflect that this is a sample. Finally, to get adjusted percentiles, one performs the following subtraction for all percentiles over 50: percentile = 100 – percentile. One can then obtain the point estimate for the total analytical error interval by taking the differences associated with the desired percentiles. For example, one would take differences associated with the 2.5^{th} and 97.5^{th} percentiles to get the total analytical error interval to contain 95% of the differences.

A two-sided nonparametric 100(1-α)% tolerance interval to contain at least 100p% of the sampled differences from a sample of size n is $[x_{(l)}, x_{(u)}]$. The appropriate values of l and u depend on α, p, and n and can be derived from published tables or figures,[10] or derived via computer code using an integrated beta density function. Appendix B gives a number used to compute the total number of observations that need to be *removed* from the sample of ordered differences. For two-sided intervals, the table value

minus 2 is divided by 2 to get the number of differences that should be removed from each end of the list of ordered differences. When the sample size is odd, the number removed from one end is rounded down to the next integer and the number removed from the other end is rounded up. That is, when the sample size is odd, there are two possible (equally correct) tolerance intervals.

6.5 Parametric Analysis

The parametric analysis assumes a normal distribution for the population of differences.

One must first assess the distribution of differences, which should be normal. This can be done visually be preparing a normal probability plot. Non-normal data will not have a bell shape. Although one can perform statistical tests for normality, this is somewhat beyond the scope of this guideline, as is the transformation procedure that is required if the data are not normal. If one has doubt that the data are normal, one should use the nonparametric estimates.

If the data are normal, total analytical error is estimated as follows. One calculates the average difference ($\bar{x}$) and its standard deviation (s) of the differences.

Total analytical error = $\bar{x} \pm (t \cdot s)$

where t is a factor from a t-distribution corresponding to the degrees of freedom (the number of differences minus one) and the specified proportion of the distribution of differences. This range is the point estimate.
The calculation of a two-sided 100(1-α)% tolerance interval to contain at least 100p% of the sampled differences from a sample of size n is

$\bar{x} \pm (k \cdot s)$

where k depends on α, p, and n. The value of k can be obtained from published tables or figures,[11,12] or derived via computer code using the noncentral t-distribution. A table of factors is contained in Appendix A.

6.6 A Review of Modeling Approaches

A textbook method[13] for estimating total analytical error, based on Westgard,[1] involves adding a multiple of precision to average bias. This is equivalent to dealing with the errors in the cause and effect diagram called "Total Imprecision," "Constant Bias" and "Linear Bias" (also called "Proportional Bias"). If any of the other error boxes have an effect on the assay, this method will underestimate the total analytical error.[4] This method was also used to model total analytical error on cholesterol[14] and glucose.[15] The glucose modeling method was critiqued by Krouwer.[16]

Another textbook method,[17] based on work by Lawton, et al.,[18] improves on this situation by accounting for the "Random Interferences" error box.

A method[19] deals with the "Protocol-Specific Bias" box in addition to the "Protocol-Independent Bias" and "Total Imprecision" boxes.

The basic problem with all of these methods is that they require all effects that might occur to be specified and estimated, as well as a calculation method for their combination into total analytical error. Since these are parametric methods, distributional assumptions are also required, as well as estimates of distribution parameters and a simulation program.

To verify that answers are correct, enough data must be collected so that a direct estimation of total analytical error is used to compare the results. If the results agree, then there is evidence that the modeling method is appropriate.

The level of effort required for these methods is huge compared to the methods recommended in Sections 6.4 and 6.5. A modeling method might nevertheless be appropriate (e.g., for a manufacturer) when one wants detailed knowledge of all possible error sources and how they affect total analytical error.

6.7 An Outline of a Simulation Method to Estimate Total Analytical Error

The steps required for a simulation method to estimate total analytical error follow. First, it is helpful to prepare block diagrams and/or flow charts of the instrument and reagent systems. One must also decide what level of error detail will be assessed and prepare a computer program representation of the assay. As one example, an assay response is transformed into the assay concentration result by a calibration equation, so part of this representation contains the calibration equation. Error sources for each step in the process must be added. Consider calibration value assignment as one such error. A manufacturer prepares a calibrator and assigns a concentration value to the vial. The analyzer software uses this value in the calculation of the calibration equation. Yet, as in most processes, there will be a difference between the actual amount of analyte in the vial and the value stated on the vial. This difference will cause an error in every result assayed for that calibrator lot (although in theory assay imprecision during calibration might combine with the calibration value assignment error to cancel both errors). To know how much error to add for calibration value assignment in the simulation, one must assess the differences between the actual amount of analyte in the vial and the value stated on the vial. This can be performed with experiments on retainments from calibrator lots. This observed distribution of differences is what is used in the simulation, so that each time a calibrator lot is selected during the simulation, a random amount of error is selected from this error distribution. This process is repeated for all error sources.

6.8 Comparison of the Total Analytical Error Evaluation Protocol With Other Evaluation Methods

Most evaluation protocols assess a particular error source. For example, NCCLS document EP6—*Evaluation of the Linearity of Quantitative Analytical Methods* assesses nonlinearity and EP7—*Interference Testing in Clinical Chemistry* assesses interferences. If one suspects a specific error source to be present, use of protocols (such as EP6 and EP7) is the most efficient way of determining the importance of that error source.

The total analytical error protocol is different in that one need not make any assumptions about error sources. If an error source is present in the specimens or instrument system, it will contribute to the total analytical error observed *regardless* of whether one knows about the error source or not. This is an advantage of a total analytical error protocol, since one may not think of all of the possible error sources. Use of the total analytical error protocol does not preclude use of any other evaluation protocols.

7 Presentation of Results with Examples

7.1 Result Report Format

Total analytical error results should be reported either as differences or as percent differences and compared to the total analytical error goal. The report should include the specifics of the experimental design, such as the number of samples, which instrument was run first for each sample, the time interval between sampling on each instrument, and so forth.

As an example, for a glucose assay, 95% of the differences were within –3 and +4 mg/dL for the concentration range tested, which was 50 to 210 mg/dL. This meets the goal of having 95% of the differences within ±5 mg/dL for this laboratory.

7.2 Result Report Interpretation

If results are well within the total analytical error goal, then one has found no evidence that the assay has failed the requirement for total analytical error. This rather weak conclusion stems from the fact that the experiment is a very small sample of the population. The issue with a small sample is that the population is complex, since each patient sample represents a unique mixture of substances.

If results are just within a 95% total analytical error goal, then the same conclusion as above applies. However, one must be cautious, since it is implied that about 5% of the patient samples are beyond the total analytical error goal. If 30 samples per day are assayed, then this represents over 500 patient samples per year that are outside of the total analytical error goal. By looking at the mountain plot, one can assess how far these samples are from the total analytical error goal.

7.3 Example 1. LDL Cholesterol

7.3.1 Data Collection

In the example, 100 randomly collected patient samples were collected and assayed by a reference and candidate method for LDL cholesterol. Each sample was assayed in duplicate for each method. The mean of the reference method was calculated. Since only single samples were to be assayed in routine use, only one of the replicates was used for the candidate method. The second replicate (not shown) was examined to ensure that no mistakes had been made. All of the data was used in this example, even though there were several discrepant results.[20]

7.3.2 Initial Calculations

Calculations can be (and were) performed in a spreadsheet. The results are displayed in Table 2. Because the comparison assay was a reference assay, the difference between the candidate result and the reference assay was used rather than the difference between the candidate result and the average between the candidate assay and reference assay.

Column A contains the number of observations. Note that Columns B through D have been affected by sorting (see Column D).

Column B contains the candidate assay results.

Column C contains the reference assay results. In this case, the result shown is the average of duplicates.

Column D contains the difference between Columns B and C.

Using the sort function in the spreadsheet application, Column D now contains the sorted difference.

Column E contains the result of the spreadsheet rank function on Column D.

Column F contains the result of the percentile (cumulative probability) whereby percentile = Column E/(N+1), where N is the total number of samples assayed, which is 100, in this example.

Column G is the folded percentile = Column F for all percentiles <0.5, and 1-Column F for all percentiles ≥ 0.5.

7.3.3 Graphical Analysis

Two plots were constructed from the data. The difference plot depicted in Figure 2 is a scatter plot of Column D (y-axis) versus Column B (x-axis).

The mountain plot depicted in Figure 3 is a scatter plot of Column G (y-axis) versus Column D (x-axis).

To facilitate interpretation of the plots, the following lines were added. Since the total analytical error goal for LDL cholesterol is ±10 mg/dL, horizontal lines at ±10 mg/dL were drawn on the difference plot. For the mountain plot, vertical lines were drawn at ±10 mg/dL and a horizontal line was drawn at the 2.5th percentile.

7.3.4 A More Formal Analysis

The differences did not appear to be normally distributed (see mountain plot); normal distribution is required to estimate total analytical error by the Bland and Altman method.[3] It appeared that the three highest differences (99.5, 131, and 219) were responsible for the non-normality. In the Bland and Altman method, one computes the standard deviation of the differences, multiplies this number by the value from a t-table corresponding to the number of degrees of freedom (total number of samples assayed – 1), and adds and subtracts this value to the mean difference. To provide further evidence that this data set was not appropriate for this type of analysis, the standard deviation of the differences was computed for all of the data (14.7) and for all of the data except for the three largest differences (6.2). Eliminating three values would thus cause the estimate to be cut in half.

Thus, the nonparametric method was used. One has to interpolate values to obtain differences corresponding to the 2.5th and 97.5th percentiles. This gave estimates of –16.8 and +115.3.

To calculate the tolerance interval corresponding to α=0.05, p=0.95, and n=100, refer to Appendix B. From this table, the total number of points to be removed is 2-2=0. Therefore, no differences should be removed from each end of the ordered list.

The 95% tolerance interval to contain at least 95% of the differences is $[x_{(1)},x_{(100)}]$ = [-18, 219].

7.3.5 Interpretation

Clearly, this assay did not meet its total analytical error goal of ±10 mg/dL. Yet only three samples were responsible for this assay showing such poor performance. This would certainly prompt this laboratory to re-examine the results. Was there a possibility of misidentified specimens? Could this have been an error with the reference assay? Note that the large total analytical error limits found suggest that the relative contribution to total analytical error by imprecision in the reference method may be small and unimportant (see also Section 4.4).

Given these results, it is up to the user to decide how to proceed. The high bias, while infrequent, would be likely to influence treatment decisions, such as statin therapy. On the other hand, the candidate assay might nevertheless perform better than other assays.

Table 2. Raw Data with All Calculations

A	B	C	D	E	F	G
Observation	Y	X	Differences	Rank	Cumulative Probability	Adjusted Cumulative Probability
1	74	92	-18	1	0.010	0.010
2	94	111	-17	2	0.020	0.020
3	101	117.5	-16.5	3	0.030	0.030
4	92	107	-15	4	0.040	0.040
5	90	105	-15	4	0.040	0.040
6	205	219.5	-14.5	6	0.059	0.059
7	50	64.5	-14.5	6	0.059	0.059
8	126	140	-14	8	0.079	0.079
9	135	149	-14	8	0.079	0.079
10	111	124.5	-13.5	10	0.099	0.099
11	131	144	-13	11	0.109	0.109
12	104	117	-13	11	0.109	0.109
13	139	151	-12	13	0.129	0.129
14	153	164.5	-11.5	14	0.139	0.139
15	94	104	-10	15	0.149	0.149
16	107	116.5	-9.5	16	0.158	0.158
17	99	108	-9	17	0.168	0.168
18	176	184.5	-8.5	18	0.178	0.178
19	123	131.5	-8.5	18	0.178	0.178
20	39	47	-8	20	0.198	0.198
21	43	51	-8	20	0.198	0.198
22	166	173.5	-7.5	22	0.218	0.218
23	100	107.5	-7.5	22	0.218	0.218
24	98	104.5	-6.5	24	0.238	0.238
25	106	112	-6	25	0.248	0.248
26	97	103	-6	25	0.248	0.248
27	123	129	-6	25	0.248	0.248
28	83	88	-5	28	0.277	0.277
29	94	99	-5	28	0.277	0.277
30	105	109.5	-4.5	30	0.297	0.297
31	122	126	-4	31	0.307	0.307
32	108	112	-4	31	0.307	0.307
33	145	148.5	-3.5	33	0.327	0.327
34	56	59.5	-3.5	33	0.327	0.327
35	137	140.5	-3.5	33	0.327	0.327
36	75	78	-3	36	0.356	0.356
37	190	192	-2	37	0.366	0.366
38	89	91	-2	37	0.366	0.366
39	133	134.5	-1.5	39	0.386	0.386
40	131	132	-1	40	0.396	0.396
41	96	97	-1	40	0.396	0.396
42	101	102	-1	40	0.396	0.396
43	174	174.5	-0.5	43	0.426	0.426
44	99	99	0	44	0.436	0.436
45	103	102	1	45	0.446	0.446

Table 2. (Continued)

A	B	C	D	E	F	G
Observation	Y	X	Differences	Rank	Cumulative Probability	Adjusted Cumulative Probability
46	115	114	1	45	0.446	0.446
47	105	104	1	45	0.446	0.446
48	102	100.5	1.5	48	0.475	0.475
49	116	114.5	1.5	48	0.475	0.475
50	128	126	2	50	0.495	0.495
51	145	143	2	50	0.495	0.495
52	155	153	2	50	0.495	0.495
53	103	101	2	50	0.495	0.495
54	179	177	2	50	0.495	0.495
55	75	73	2	50	0.495	0.495
56	169	166.5	2.5	56	0.554	0.446
57	155	151.5	3.5	57	0.564	0.436
58	72	68.5	3.5	57	0.564	0.436
59	127	123.5	3.5	57	0.564	0.436
60	133	129	4	60	0.594	0.406
61	76	71.5	4.5	61	0.604	0.396
62	116	111	5	62	0.614	0.386
63	103	98	5	62	0.614	0.386
64	224	219	5	62	0.614	0.386
65	135	129	6	65	0.644	0.356
66	99	93	6	65	0.644	0.356
67	120	113	7	67	0.663	0.337
68	123	116	7	67	0.663	0.337
69	118	111	7	67	0.663	0.337
70	119	111.5	7.5	70	0.693	0.307
71	132	123	9	71	0.703	0.297
72	136	127	9	71	0.703	0.297
73	121	112	9	71	0.703	0.297
74	116	106.5	9.5	74	0.733	0.267
75	84	74	10	75	0.743	0.257
76	156	146	10	75	0.743	0.257
77	197	187	10	75	0.743	0.257
78	110	99.5	10.5	78	0.772	0.228
79	87	76.5	10.5	78	0.772	0.228
80	172	160.5	11.5	80	0.792	0.208
81	236	224.5	11.5	80	0.792	0.208
82	221	209	12	82	0.812	0.188
83	133	120	13	83	0.822	0.178
84	105	92	13	83	0.822	0.178
85	116	102	14	85	0.842	0.158
86	123	108.5	14.5	86	0.851	0.149
87	113	98	15	87	0.861	0.139
88	178	159	19	88	0.871	0.129
89	180	161	19	88	0.871	0.129
90	175	154.5	20.5	90	0.891	0.109

Table 2. (Continued)

A	B	C	D	E	F	G
Observation	**Y**	**X**	**Differences**	**Rank**	**Cumulative Probability**	**Adjusted Cumulative Probability**
91	198	174	24	91	0.901	0.099
92	144	118	26	92	0.911	0.089
93	160	134	26	92	0.911	0.089
94	204	177	27	94	0.931	0.069
95	131	102.5	28.5	95	0.941	0.059
96	142	102.5	39.5	96	0.950	0.050
97	181	131.5	49.5	97	0.960	0.040
98	154	54.5	99.5	98	0.970	0.030
99	213	82	131	99	0.980	0.020
100	284	65	219	100	0.990	0.010

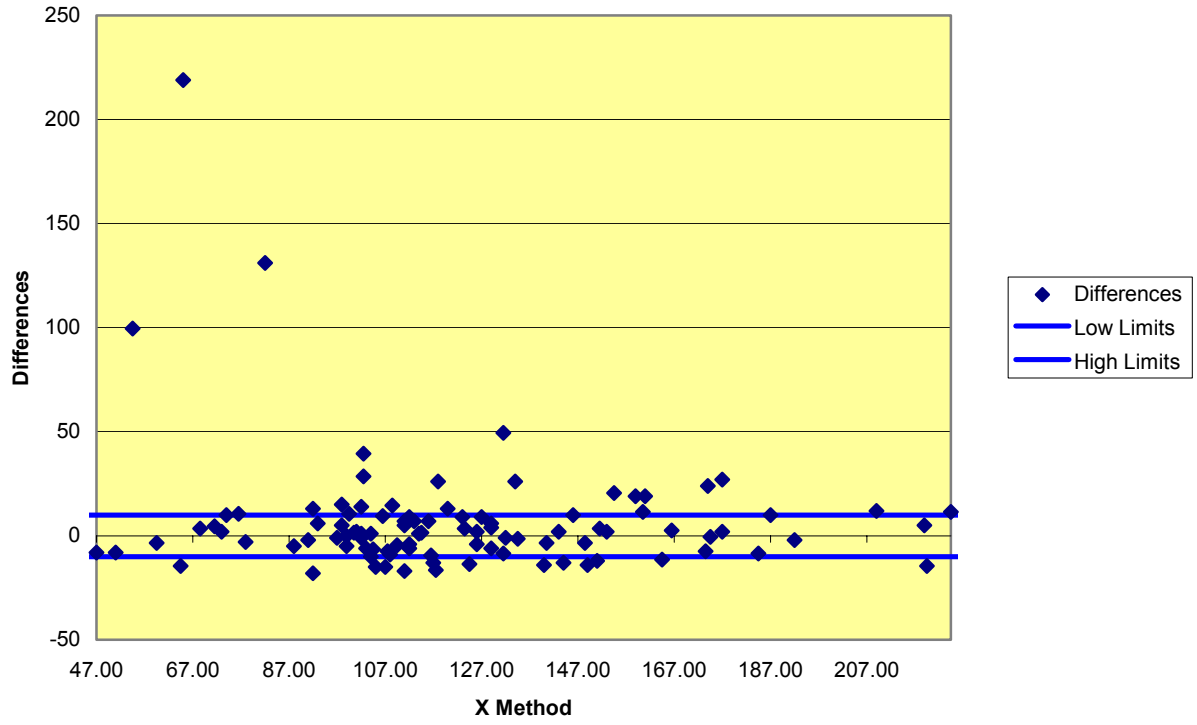

Figure 2. Difference Plot LDL Cholesterol

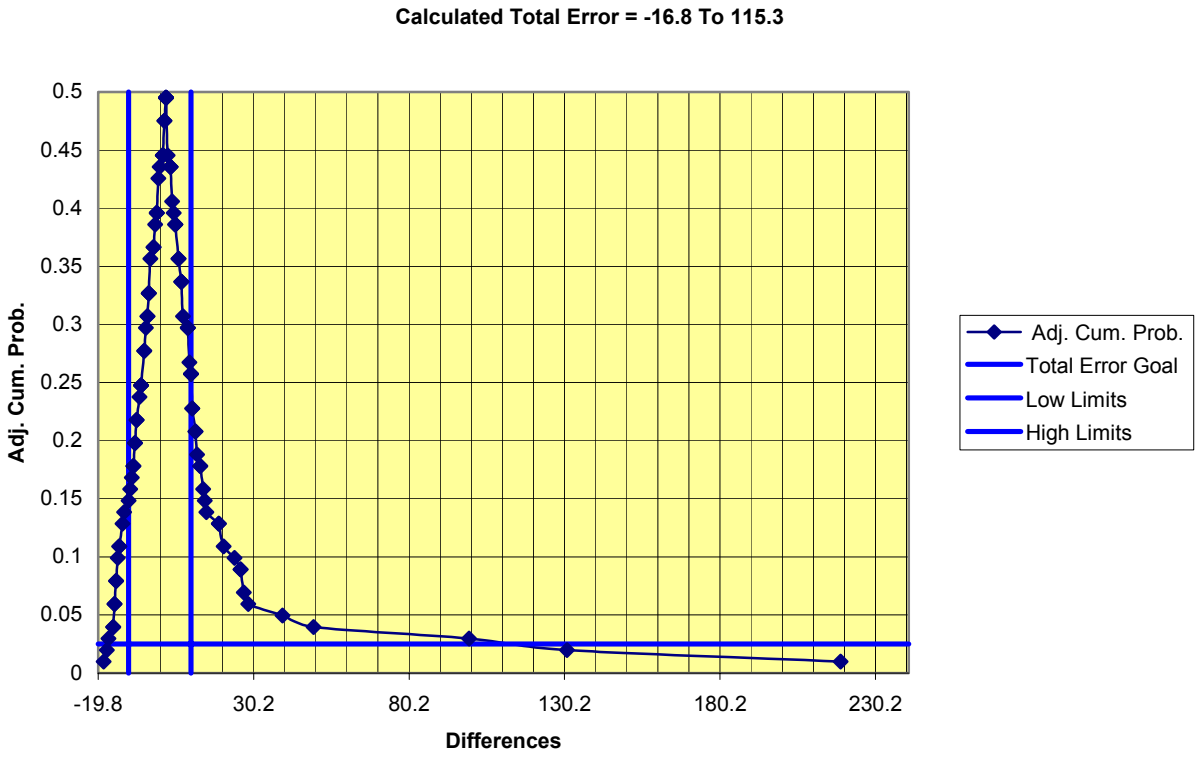

Figure 3. Mountain Plot LDL Cholesterol

7.4 Example 2. Sodium

7.4.1 Data Collection

In the example, 120 randomly collected patient samples were collected and assayed once by a candidate method for sodium and by the current method used in the laboratory.

7.4.2 Initial Calculations

Calculations can be (and were) performed in a spreadsheet. The results are displayed in Table 3. The difference between the candidate result and the average between the candidate and comparison assay was used.

Column A contains the number of observations. Note that Columns B through D have been affected by sorting. (See Column D.)

Column B contains the candidate assay results.

Column C contains the comparison assay results.

Column D contains the difference between Columns B and the average between Columns B and C.

Using the sort function in the spreadsheet application, Column D now contains the sorted difference.

Column E contains the result of the spreadsheet application rank function on Column D.

Column F contains the result of the percentile (cumulative probability) whereby percentile = Column E/(N+1), where N is the total number of samples assayed, which is 125, in this example.

Column G is the folded percentile = Column G for all percentiles <0.5, and 1-Column F for all percentiles ≥0.5.

7.4.3 Graphical Analysis

Two plots were constructed from the data. The difference plot depicted in Figure 4 is a scatter plot of Column D (y-axis) versus Column B (x-axis). The mountain plot depicted in Figure 5 is a scatter plot of Column G (y-axis) versus Column D (x-axis).

To facilitate interpretation of the plots, the following lines were added. Since the total analytical error goal for sodium is ±4 mmol/L, horizontal lines at ±4 mmol/L were drawn on the difference plot. For the mountain plot, vertical lines were drawn at ±4 mmol/L and a horizontal line was drawn at the 2.5th percentile.

7.4.4 A More Formal Analysis

The differences appear to be normally distributed (see difference plot); normal distribution is required to estimate total analytical error by the Bland and Altman method. In the Bland and Altman method, one computes the standard deviation of the differences, multiplies this number by the value from a t-table corresponding to the number of degrees of freedom (total number of samples assayed – 1), and adds and subtracts this value to the mean difference. The results are:

Average difference = -0.0832
Standard deviation of differences = 1.458

t-value multiplier = 1.979
Lower value for total analytical error = -2.97
Upper value for total analytical error = 2.80

The values: -3 and 2.8 appear on the difference plot.

For the parametric tolerance interval, k is interpolated from the table in Appendix A = 2.199.

$-0.0832 \pm (2.199 \times 1.458)$ gives -0.0832 ± 3.206 or $(-3.289, 3.123)$.

The 95% tolerance interval to contain at least 95% of the differences is (–3.29, 3.12).

For the nonparametric method, one has to interpolate values to obtain differences corresponding to the 2.5^{th} and 97.5^{th} percentiles. This gave estimates of –3.3 and +2.7, which are displayed on the mountain plot chart.

For the nonparametric tolerance interval, the total number of points to be removed is 3-2=1, based on interpolation from Appendix B. Since 1/2=0.5, then one difference should be removed from one end and none should be removed from the other. There are two different tolerance intervals that can be computed.

One 95% tolerance interval to contain at least 95% of the differences is $[x_{(2)},x_{(120)}]$ = [-3.5, 3]. Alternatively, another 95% tolerance interval to contain at least 95% of the differences is $[x_{(1)}\ x_{(124)}]$ = [-3.5, 3]

In this particular example, the two possible intervals happen to be identical.

7.4.5 Interpretation

The difference plot shows an apparent random scatter with all of the data within limits. The mountain plot is centered close to zero difference and does not show any outliers. One can conclude that this assay meets its goal of having 95% of its results differing less than ±4 mmol/L.

Table 3. Initial Resultant Calculations from Data Collection for Sodium Assays

A	B	C	D	E	F	G
Observation	Sodium Candidate	Sodium Comparison	Differences	Rank	Cumulative Probability	Adjusted Cumulative Probability
1	129.2	136.2	-3.50	1	0.008	0.008
2	129.8	136.8	-3.50	1	0.008	0.008
3	125.3	132.2	-3.45	3	0.024	0.024
4	125.5	132.4	-3.45	3	0.024	0.024
5	133.5	139.9	-3.20	5	0.040	0.040
6	144.3	149.9	-2.80	6	0.048	0.048
7	139.5	144.7	-2.60	7	0.056	0.056
8	132.9	137.3	-2.20	8	0.063	0.063
9	143.0	147.2	-2.10	9	0.071	0.071
10	135.0	139.1	-2.05	10	0.079	0.079
11	142.8	146.9	-2.05	10	0.079	0.079
12	132.0	136.0	-2.00	12	0.095	0.095
13	145.3	149.2	-1.95	13	0.103	0.103
14	143.5	147.4	-1.95	13	0.103	0.103
15	132.2	135.8	-1.80	15	0.119	0.119
16	143.4	147.0	-1.80	15	0.119	0.119
17	147.4	150.9	-1.75	17	0.135	0.135
18	136.8	140.2	-1.70	18	0.143	0.143
19	145.3	148.7	-1.70	18	0.143	0.143
20	143.6	146.8	-1.60	20	0.159	0.159
21	146.3	149.5	-1.60	20	0.159	0.159
22	127.1	130.0	-1.45	22	0.175	0.175
23	129.2	132.1	-1.45	22	0.175	0.175
24	136.6	139.4	-1.40	24	0.190	0.190
25	142.7	145.4	-1.35	25	0.198	0.198
26	148.2	150.7	-1.25	26	0.206	0.206
27	133.2	135.7	-1.25	26	0.206	0.206
28	140.4	142.9	-1.25	26	0.206	0.206
29	143.4	145.9	-1.25	26	0.206	0.206
30	133.1	135.6	-1.25	26	0.206	0.206
31	133.9	135.9	-1.00	31	0.246	0.246
32	140.8	142.8	-1.00	31	0.246	0.246
33	148.1	150.1	-1.00	31	0.246	0.246
34	140.6	142.5	-0.95	34	0.270	0.270
35	128.2	130.1	-0.95	34	0.270	0.270
36	130.6	132.4	-0.90	36	0.286	0.286
37	146.2	147.9	-0.85	37	0.294	0.294
38	141.6	143.3	-0.85	37	0.294	0.294
39	136.3	138.0	-0.85	37	0.294	0.294
40	144.0	145.5	-0.75	40	0.317	0.317
41	136.1	137.6	-0.75	40	0.317	0.317
42	137.7	139.1	-0.70	42	0.333	0.333

Table 3. (Continued)

A	B	C	D	E	F	G
Observation	Sodium Candidate	Sodium Comparison	Differences	Rank	Cumulative Probability	Adjusted Cumulative Probability
43	148.3	149.7	-0.70	42	0.333	0.333
44	136.9	138.2	-0.65	44	0.349	0.349
45	125.7	126.9	-0.60	45	0.357	0.357
46	124.7	125.8	-0.55	46	0.365	0.365
47	137.6	138.7	-0.55	46	0.365	0.365
48	141.1	142.0	-0.45	48	0.381	0.381
49	132.5	133.4	-0.45	48	0.381	0.381
50	141.7	142.6	-0.45	48	0.381	0.381
51	145.4	146.2	-0.40	51	0.405	0.405
52	142.2	143.0	-0.40	51	0.405	0.405
53	136.3	137.0	-0.35	53	0.421	0.421
54	143.6	144.1	-0.25	54	0.429	0.429
55	141.7	142.1	-0.20	55	0.437	0.437
56	139.3	139.7	-0.20	55	0.437	0.437
57	145.7	146.0	-0.15	57	0.452	0.452
58	144.6	144.8	-0.10	58	0.460	0.460
59	141.5	141.7	-0.10	58	0.460	0.460
60	147.7	147.8	-0.05	60	0.476	0.476
61	132.1	132.1	0.00	61	0.484	0.484
62	141.6	141.5	0.05	62	0.492	0.492
63	145.6	145.5	0.05	62	0.492	0.492
64	137.6	137.5	0.05	62	0.492	0.492
65	152.0	151.9	0.05	62	0.492	0.492
66	147.5	147.4	0.05	62	0.492	0.492
67	139.8	139.7	0.05	62	0.492	0.492
68	140.6	140.4	0.10	68	0.540	0.460
69	144.1	143.8	0.15	69	0.548	0.452
70	150.4	150.1	0.15	69	0.548	0.452
71	140.2	139.8	0.20	71	0.563	0.437
72	140.1	139.7	0.20	71	0.563	0.437
73	135.6	135.1	0.25	73	0.579	0.421
74	140.6	140.0	0.30	74	0.587	0.413
75	144.2	143.6	0.30	74	0.587	0.413
76	135.0	134.4	0.30	74	0.587	0.413
77	150.2	149.6	0.30	74	0.587	0.413
78	156.4	155.7	0.35	78	0.619	0.381
79	138.1	137.3	0.40	79	0.627	0.373
80	140.8	140.0	0.40	79	0.627	0.373
81	144.0	143.1	0.45	81	0.643	0.357
82	141.9	140.9	0.50	82	0.651	0.349
83	148.8	147.8	0.50	82	0.651	0.349
84	134.7	133.7	0.50	82	0.651	0.349

Table 3. (Continued)

A	B	C	D	E	F	G
Observation	Sodium Candidate	Sodium Comparison	Differences	Rank	Cumulative Probability	Adjusted Cumulative Probability
85	143.2	142.1	0.55	85	0.675	0.325
86	141.0	139.9	0.55	85	0.675	0.325
87	159.0	157.8	0.60	87	0.690	0.310
88	137.5	136.3	0.60	87	0.690	0.310
89	137.5	136.3	0.60	87	0.690	0.310
90	144.8	143.5	0.65	90	0.714	0.286
91	142.8	141.3	0.75	91	0.722	0.278
92	143.5	141.8	0.85	92	0.730	0.270
93	148.5	146.7	0.9	93	0.738	0.262
94	148.5	146.5	1.00	94	0.746	0.254
95	148.1	146.1	1.00	94	0.746	0.254
96	147.7	145.5	1.10	96	0.762	0.238
97	144.3	142.1	1.10	96	0.762	0.238
98	142.8	140.5	1.15	98	0.778	0.222
99	149.3	146.9	1.20	99	0.786	0.214
100	144.1	141.7	1.20	99	0.786	0.214
101	140.8	138.3	1.25	101	0.802	0.198
102	134.1	131.6	1.25	101	0.802	0.198
103	147.4	144.9	1.25	101	0.802	0.198
104	140.9	138.2	1.35	104	0.825	0.175
105	133.3	130.6	1.35	104	0.825	0.175
106	140.7	137.9	1.40	106	0.841	0.159
107	138.5	135.6	1.45	107	0.849	0.151
108	148.2	145.1	1.55	108	0.857	0.143
109	144.4	141.2	1.60	109	0.865	0.135
110	132.5	129.1	1.70	110	0.873	0.127
111	148.7	145.2	1.75	111	0.881	0.119
112	145.2	141.6	1.80	112	0.889	0.111
113	148.9	145.0	1.95	113	0.897	0.103
114	165.0	161.0	2.00	114	0.905	0.095
115	132.1	128.1	2.00	114	0.905	0.095
116	160.8	156.7	2.05	116	0.921	0.079
117	140.1	135.9	2.10	117	0.929	0.071
118	144.3	140.1	2.10	117	0.929	0.071
119	158.6	154.3	2.15	119	0.944	0.056
120	164.3	159.9	2.20	120	0.952	0.048
121	151.5	146.7	2.40	121	0.960	0.040
122	142.0	136.9	2.55	122	0.968	0.032
123	141.2	135.7	2.75	123	0.976	0.024
124	150.7	144.7	3.00	124	0.984	0.016
125	157.9	151.9	3.00	124	0.984	0.016

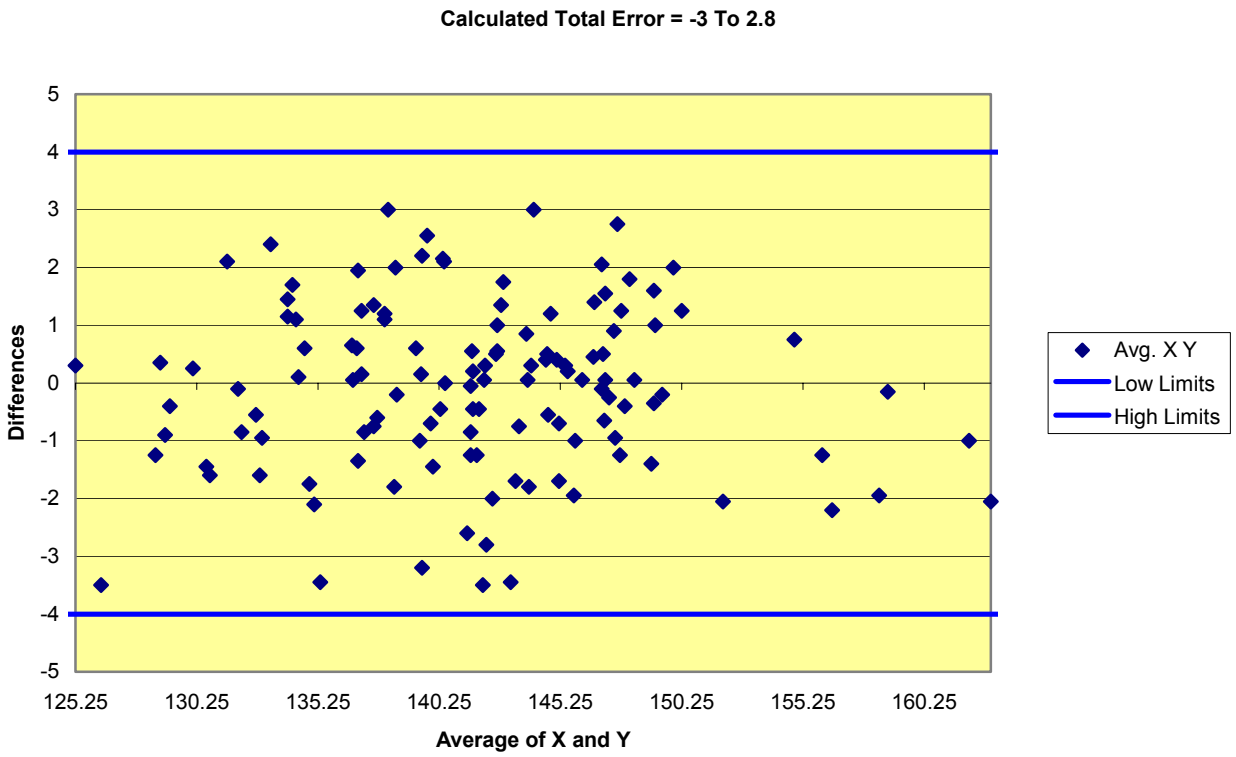

Figure 4. Difference Plot Sodium

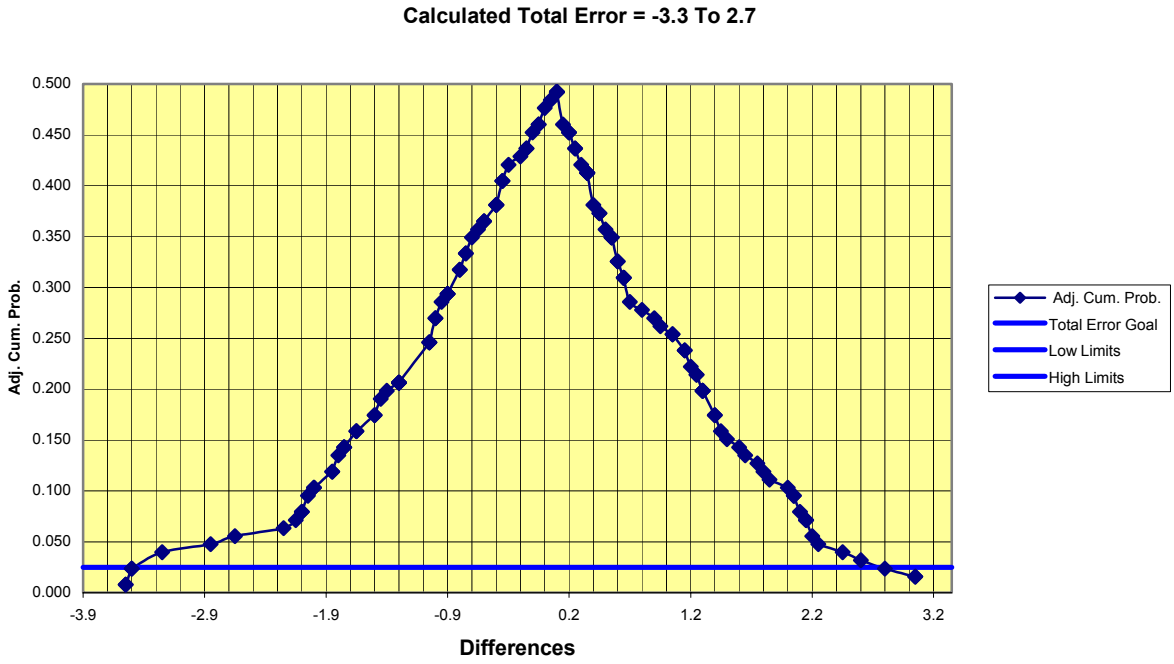

Figure 5. Mountain Plot Sodium

References

[1] Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem*. 1974;20:825-833.

[2] Eisenhart C. Realistic evaluation of the precision and accuracy of instrument calibration systems. *J Res NBS*. 1963;67C(2):161-187.

[3] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* February 1986.

[4] Krouwer JS. Estimating total analytical error and its sources – techniques to improve method evaluation. *Arch Pathol Lab Med*. 1992;116:726-731.

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

[6] Petersen PH, Fraser CG, Kallner A, Kenny D, eds. Special issue strategies to set global analytical quality specifications in laboratory medicine. *Scand J Clin Lab Invest.* 1999;59:475-585.

[7] International Organization for Standardization (ISO). *Determination of Analytical Performance Goals for Laboratory Procedures Based on Medical Requirements.* ISO 15196. Geneva: 2002.

[8] Clark WL, Cox D, Gonder-Frederick LA, et al. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care.* 1987;10:622-628.

[9] Krouwer JS, Monti KL. A simple, graphical method to evaluate laboratory assays – short communication. *Euro J Clin Chem Clin Biochem*. 1995;33:525-527.

[10] Monti KL. Folded empirical distribution function curves-mountain plots. *The American Statistician*. 1995;49:342-345.

[11] Hahn GJ, Meeker WQ. *Statistical Intervals: A Guide for Practitioners.* New York: John Wiley and Sons; 1991.

[12] Odeh RE, Owen DB. *Tables for Normal Tolerance Limits, Sampling Plans, and Screening.* New York: Marcel Dekker, Inc.; 1980.

[13] Carey RN, Garber CC. Evaluation of methods. In: Kaplan LA, Pesce AJ, eds. *Clinical Chemistry Theory, Analysis, and Correlation.* St Louis: Mosby; 1989.

[14] U.S. Dept. of Health and Human Services. *Recommendations for Improving Cholesterol Measurements*. NIH 90-2964. Washington, D.C.: NIH; 1990.

[15] Boyd JC, Bruns DE. Quality specifications for glucose meters; assessment by simulation modeling of errors in insulin dose. *Clin Chem.* 2001;47:209-214.

[16] Krouwer JS. How to improve total analytical error modeling by accounting for error sources beyond imprecision and bias. *Clin Chem.* 2001;47;1329-1330.

[17] Kringle RO, Johnson GF. Acquisition, management, and application of laboratory data – statistical procedures. In: Tietz NW, ed. *Textbook of Clinical Chemistry*. Philadelphia: W.B. Saunders; 1986.

[18] Lawton WH, Sylvester EA, Young-Ferraro BJ. Statistical comparison of multiple analytic procedures: application to clinical chemistry. *Technometrics*. 1979;21:397-409.

[19] Krouwer JS. Multi-factor designs IV: How multifactor designs improve the estimate of total analytical error by accounting for protocol specific biases. *Clin Chem*. 1991;37:26-29.

[20] Miller WG, Maymack PD, Anderson FP, Ethridge SF, Jayne EC. Performance of Four Homogeneous Direct Methods for LDL-Cholestrerol. *Clin Chem*. 2002; 48:489-498.

Appendix A. Factors (k) to Calculate Normal Distribution Tolerance Intervals

Below is a table of factors (k) to calculate normal distribution two-sided 100(1-α)% tolerance intervals for the center of the distribution. *Gamma* refers to the amount of the distribution to be included, (e.g., 0.95 means 95% of the distribution will be included).

gamma	0.9	0.95	0.99	0.9	0.95	0.99
(1-α)	0.9	0.9	0.9	0.95	0.95	0.95
n						
30	2.03	2.41	3.17	2.14	2.55	3.35
31	2.02	2.4	3.16	2.13	2.54	3.33
32	2.01	2.39	3.15	2.12	2.52	3.32
33	2	2.38	3.13	2.11	2.51	3.3
34	1.99	2.38	3.12	2.1	2.5	3.29
35	1.99	2.37	3.11	2.09	2.49	3.27
36	1.98	2.36	3.1	2.08	2.48	3.26
37	1.98	2.35	3.09	2.07	2.47	3.25
38	1.97	2.35	3.08	2.07	2.46	3.24
39	1.96	2.34	3.08	2.06	2.45	3.22
40	1.96	2.33	3.07	2.05	2.44	3.21
41	1.95	2.33	3.06	2.05	2.44	3.2
42	1.95	2.32	3.05	2.04	2.43	3.19
43	1.94	2.32	3.04	2.03	2.42	3.18
44	1.94	2.31	3.04	2.03	2.42	3.17
45	1.94	2.31	3.03	2.02	2.41	3.17
46	1.93	2.3	3.02	2.02	2.4	3.16
47	1.93	2.3	3.02	2.01	2.4	3.15
48	1.92	2.29	3.01	2.01	2.39	3.14
49	1.92	2.29	3.01	2	2.38	3.13
50	1.92	2.28	3	2	2.38	3.13
55	1.9	2.26	2.98	1.98	2.35	3.09
60	1.89	2.25	2.96	1.96	2.33	3.07
65	1.88	2.23	2.94	1.94	2.31	3.04
70	1.87	2.22	2.92	1.93	2.3	3.02
75	1.86	2.21	2.91	1.92	2.28	3
80	1.85	2.2	2.89	1.91	2.27	2.99
85	1.84	2.19	2.88	1.9	2.26	2.97
90	1.83	2.19	2.87	1.89	2.25	2.96
95	1.83	2.18	2.86	1.88	2.24	2.95
100	1.82	2.17	2.85	1.87	2.23	2.93
120	1.8	2.15	2.83	1.85	2.2	2.9
140	1.79	2.13	2.8	1.83	2.18	2.87
160	1.78	2.12	2.79	1.82	2.17	2.85
180	1.77	2.11	2.77	1.81	2.15	2.83
200	1.76	2.1	2.76	1.8	2.14	2.82
500	1.72	2.05	2.69	1.74	2.07	2.72
1000	1.69	2.02	2.65	1.71	2.04	2.68
Infinity	1.64	1.96	2.58	1.64	1.96	2.58

Appendix B. Number of Extreme Observations ν to Be Removed from the Ends of a Sample Size *n* to Obtain a Two-Sided Distribution-Free Tolerance Interval or to Obtain a One-Sided Distribution-Free Tolerance Bound That Contains at Least 100*p*% of the Sample Population with 100(1-α)% Confidence

(From Hahn GJ, Meeker WQ. *Statistical Intervals: A Guide for Practitioners.* © 1991 John Wiley and Sons, Inc. This material is used by permission from John Wiley and Sons, Inc., and the Institute of Mathematical Statistics.)

		p=0.900			*p*=0.950			*p*=0.990		
n	1-α:	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
10		1	1	1	1	1	1	1	1	1
		0.6513*	0.6513*	0.6513*	0.4013*	0.4013*	0.4013*	0.0956*	0.0956*	0.0956*
15		1	1	1	1	1	1	1	1	1
		0.7941*	0.7941*	0.7941*	0.5367*	0.5367*	0.5367*	0.1399*	0.1399*	0.1399*
20		1	1	1	1	1	1	1	1	1
		0.8784*	0.8784*	0.8784*	0.6415*	0.6415*	0.6415*	0.1821*	0.1821*	0.1821*
25		1	1	1	1	1	1	1	1	1
		0.9282	0.9282*	0.9282*	0.7226*	0.7226*	0.7226*	0.2222*	0.2222*	0.2222*
30		1	1	1	1	1	1	1	1	1
		0.9576	0.9576	0.9576*	0.7854*	07854*	07854*	0.2603*	0.2603*	0.2603*
35		1	1	1	1	1	1	1	1	1
		0.9750	0.9750	0.9750*	0.8339*	0.8339*	0.8339*	0.2966*	0.2966*	0.2966*
40		2	1	1	1	1	1	1	1	1
		0.9195	0.9852	0.9852*	0.8715*	0.8715*	0.8715*	0.3310*	0.3310*	0.3310*
50		2	2	1	1	1	1	1	1	1
		0.9662	0.9662	0.9948	0.9231*	0.9231*	0.9231*	0.3950*	0.3950*	0.3950*
60		3	2	1	1	1	1	1	1	1
		0.9470	0.9862	0.9982	0.9539	0.9539	0.9539*	0.4528*	0.4528*	0.4582*
80		5	4	2	2	1	1	1	1	1
		0.9120	0.9647	0.9978	0.9139	0.9835	0.9835*	0.5225*	0.5525*	0.5525*
100		6	5	4	2	2	1	1	1	1
		0.9424	0.9763	0.9922	0.9629	0.9629	0.9941	0.6340*	0.6340*	0.6340*
200		15	13	11	6	5	4	1	1	1
		0.9071	0.9680	0.9919	0.9377	0.9736	0.9910	0.8660*	0.8660*	0.8660*
300		23	22	19	10	9	7	1	1	1
		0.9301	0.9542	0.9903	0.9350	0.9659	0.9934	0.9510*	0.9510*	0.9510*
400		32	30	27	15	13	11	2	1	1
		0.9254	09643	0.9908	0.9010	0.9645	0.9906	0.9095	0.9820	0.9820*
500		41	39	35	19	17	14	2	2	1
		0.9249	0.9607	09921	0.9135	0.9657	0.9945	0.9602	0.9602	0.9934
600		51	48	44	23	21	18	3	2	1
		0.9043	0.9591	0.9901	0.9247	0.9680	0.9938	0.9389	0.9830	0.9976
800		69	66	61	32	30	26	5	4	2
		0.9146	0.9593	0.9912	0.9199	0.9606	0.9935	0.9015	0.9583	0.9971
1000		88	85	79	41	39	35	6	5	3
		0.9081	0.9515	0.9901	0.9194	0.9566	0.9907	0.9339	0.9713	0.9973

*Indicates that a symmetric confidence interval or bound with the desired confidence level cannot be achieved.

Appendix C. A Mathematical Representation of Total Analytical Error

Mandel[1] showed how a difference between a test and reference result could be described as a combination of random and systematic error (Equation 1 below).

$$TAE = (y - R) = (y - \mu) + (\mu - R) \quad (1)$$

Equation 2 is an expansion of Equation 1 to account for n replicates of each of m different specimens.

$$TAE = \sum_{j=1}^{m}\sum_{i=1}^{n}(y_{ij} - R_j) = \sum_{j=1}^{m}\sum_{i=1}^{n}(y_{ij} - \overline{\mu_j}) + \sum_{j=1}^{m}(\overline{\mu_j} - R_j) \quad (2)$$

where

TAE is total analytical error (measurement error)

y_{ij} is the i^{th} observation from the j^{th} sample of the new method;

R_j is reference method result for the j^{th} sample; and

$\overline{\mu_j}$ is the mean of the j^{th} sample.

In Equation 2, the second double summation term is a measure of imprecision and the last term represents the distribution of bias that is observed in each sample.

Reference for Appendix C

1 Mandel J. *The Statistical Analysis of Experimental Data*. Mineola, NY: Dover Publications. 1964:104-105.

NCCLS consensus procedures include an appeals process that is described in detail in Section 9 of the Administrative Procedures. For further information, contact the Executive Offices or visit our website at www.nccls.org.

Summary of Consensus Comments and Committee Responses

EP21-P: *Estimation of Total Analytical Error for Clinical Laboratory Methods; Proposed Guideline*

General

1. The concept of total error reporting is very useful for clinical laboratories and manufacturers alike. The explanations regarding total error are very informative and well stated in the document. The protocol section, however, does not contain adequate instructions for use by manufacturers and does not address the statistics that would be available for use by the laboratory for calculating total error.

- **The only protocol difference for manufacturers is an increased sample size. Sections 6.4 and 6.5 present statistics and Section 7 provides two examples.**

2. The concept of error budgeting would be helpful for manufacturers, as would a section included in the document. Perhaps two separate sections on protocols, one for manufacturers and one for clinical labs, would be appropriate. For manufacturers, the inclusion of simulation studies would be helpful, and a more robust approach to precision testing is necessary.

- **See Comment 1. A complete treatment of error budgeting and simulation studies is beyond the scope of the document; however, Section 6.7 outlining the principles for a simulation study has been added.**

3. It would be useful to the clinical laboratory for package inserts (instructions for use) to contain information on total error. This would be a major shift in focus from separate precision and accuracy claims, but well worth the effort to move the industry toward better quality control.

- **The subcommittee agrees.**

4. I am skeptical of how well the proposed approach would adequately reflect the contribution of method reproducibility to total analytical error (TAE), relative to that afforded by the older modeling approach.

- **Appendix C has been added to describe the theory behind total analytical error.**

5. The proposed TAE analysis seems capable of yielding unfair results in cases where a new test method has significantly improved reproducibility relative to the comparison method. Some counsel to readers might be appropriate for interpretation of the results in such cases.

- **This is a valid concern and is already mentioned in three places in the document. Replication of the comparison method and use of the average value is used to minimize differences due to imprecision of the comparison method.**

6. Additional discussion is needed on patient specimen selection, especially for heteroscedastic methods. Unless care is taken to segment the concentration range and run the protocol separately for each segment, the resulting TAE could be grossly over- or underestimated.

- **This information is included in Section 5.1.**

7. EP21-P does not provide sufficient criteria for use of the nonparametric versus parametric approach, or even total versus specific errors. For example, it would seem appropriate to test for the difference between two analytical systems using total error, when the only difference between them is a new shipment of the same lot of reagent, or a probe has been changed. Use of total error versus specific error would not be appropriate when a new calibrator is being introduced or a new method, differing by at least two variables, is being tested. This document should have discussed these issues in detail.

- **The two examples provide guidance for use of the parametric versus the nonparametric method. Their use is also discussed in Section 6.2. The last paragraph of the Foreword provides general recommendations for the performance of a total analytical error study. The specific guidance suggested in this comment is beyond the scope of the document.**

Section 3

8. The definition of "frequency" uses the term "class," which was not previously defined.

- **The definition of "frequency" is not needed and has been deleted.**

9. I question the utility of including such a large set of definitions, particularly as a number of them are not even used (or just in passing) within the document. Examples of these are "analysis," "correction," "diagnostic test," "in vitro diagnostic," "qualitative," "test," and "validation." In fact, going by the definitions of Section 3, the word "sample" is used incorrectly in several cases where "specimen" is meant. Examples of these are Section 5.1 (". . . 40 patient samples should be run. . .") and Section 5.2 (". . . sufficient patient samples"). I recommend giving just the few key terms that are introduced in the document and referring to the NRSCL8 document for others.

- **Many definitions have been deleted. The recommended changes regarding specimen have been incorporated.**

Section 4.1

10. The discussion of modeling approaches to TAE states that the earlier method of Westgard only addresses components of total imprecision, constant bias, and proportional bias in the Figure l error budget. This is not correct. Performing a method comparison experiment over several days to estimate systematic error, per usual practice and as recommend by EP9-A, certainly can incorporate contributions from drift, sample carry-over, reagent carry-over, and matrix interferences (depending on specimen selection).

- **If the sampling sequence recommended in EP9-A is utilized, the effect of linear drift will be eliminated in estimating bias. However, if one were not to use the recommended sampling sequence and the error sources mentioned by the commenter were present, these error sources would affect the magnitude of proportional and constant error estimated. In this sense, the commenter is correct. These error sources play a role in the bias estimate. However, the model referred to by the commenter does not correctly account for these error sources' contribution to total analytical error.**

11. I disagree with the decision to exclude outliers as a contributor to the total analytical error (TAE) box in Figure 1 and to largely omit discussion of them within this document. To the extent that they do not reflect pre-/postanalytical errors, outliers fall within the domain of analytical error and readers should receive some basic counseling on what to do about them. While elimination of values on a solely statistical basis could underestimate the TAE, blind retention of suspect values could equally overestimate the error. Since this document relies heavily on EP9-A, I suggest a compromise might be to follow the guidance in Section 4.4 of EP9-A on use of an outlier test.

- **The commenter is correct to point out a problem with Figure 1. Section 6.3 has been expanded for clarification. Outliers can be removed when estimating error sources such as drift or proportional error (EP9), but not when estimating total analytical error. Yet, total analytical error is specified as covering less than 100% of the population so that if outliers occur, total analytical error may still be judged as acceptable. This is explained in Section 6.3. The nonparametric method of estimating total analytical error is robust to outliers.**

Section 4.1

12. This section refers to the errors in Figure 1 as medical errors when in reality they are analytical errors. Furthermore, the references to protocol-independent and protocol-specific errors, as terminology, do not make sense.

- **The top-level event (caused by a variety of events, most of which are listed as analytical) for this cause and effect diagram is morbidity/mortality due to a medical error, hence the reference to "medical errors." The discussion of protocol-independent and protocol-specific errors is not essential to the document. Users can obtain more information from Reference 17.**

Section 4.3

13. While I understand the rationale given in the comment and response section for not providing sources of TAE goals in Section 4.3, I disagree with it. I would like to see at least a reference to the existing goals (e.g., CLIA), along with a caveat that they were derived from modeling approaches and may not apply exactly with the new proposed analysis approach. It would at least provide a starting point for user considerations.

- **A reference (see reference 7) that provides a comprehensive discussion of approaches has been added.**

14. This section rejects the total analytical error goals that currently are in the literature (i.e., from CLIA and CAP). A laboratory cannot establish goals; rather, an individual (in this case the laboratory director, generally a pathologist) must perform the

task of goal setting. In most cases, we would want an individual trained in clinical pathology to make these decisions, as they have both the analytical and the medical background to integrate the information.

- **Most goals from regulatory bodies relate to proficiency surveys and are based on quality control samples. This document focuses on patient samples. Errors due to random interferences will be missed using quality control samples (see Reference 16). Describing the qualifications of the individual responsible for setting goals is beyond the scope of this document.**

15. Bias and precision should have separate goals as they are frequently used differently. EP21-P attempts to combine all sources of analytical error into a single feature. However, an important strategy to adopt is separation of bias and precision errors. One can reduce the effect of precision errors by using the standard error of the mean (multiple replicates on the same sample).

- **This document does not preclude the use of individual goals (for bias and precision). However, EP21 focuses on total analytical error; hence, discussion of bias and precision are beyond the scope of the document (although their estimation in conjunction with total analytical error is briefly discussed in Section 6.6). Moreover, bias and precision are separate NCCLS evaluation protocols (EP9 and EP5, respectively).**

Section 5

16. This section assumes one can perform all necessary studies regarding error for a method by using total analytical error, i.e., one does not need to check bias separately, nor linearity of a method, nor interferences. This policy is a dangerous move, because the error for any analytical method must be broken down into each of its parts.

- **Section 6.8 states that EP21 does not preclude one from performing other studies. Detailed discussion of other studies is beyond the scope of the document. Although specific studies (such as a test for linearity) will be the best way to detect a specific error source (in this case nonlinearity), the value of a total analytical error study is that errors from *any* source will contribute to the estimated error found. An example where traditional studies underestimated total analytical error is pointed out in Reference 4.**

Section 5.1

17. The first sentence should read that EP9 is a guideline for method comparisons, not patient comparisons.

- **This change has been incorporated.**

18. What is the basis for selecting N=125 as the nominal sample size in Section 5.1? Should readers interpret this as N=125 specimens per concentration range? If so, the text should be more explicit. I question if this number of specimens is adequate to capture all sources of error as purported by the comments in Section 6.7.

- **The section has been changed. The recommended N has been changed to 120. Evaluation protocol subcommittees are always faced with the question of how many samples should be run or, more often, what is the minimum number of samples that should be run. These questions can be answered formally by specifying goals, and Type I and Type II errors. However, this raises the complexity level of the document for nonstatisticians. Therefore, the subcommittee has used its experience to arrive at a *minimum* sample size. Note, most manufacturers will assay many more samples during development and evaluation of assays.**

Section 5.2

19. This section advocates the use of quality control materials in lieu of patient samples. Quality control materials should not be used for the routine assessment of errors for laboratory methods. The reference to insufficient resources or patient samples is inappropriate, as it provides an "excuse" to hospital administrators to remove resources from the laboratory, there being another "totally acceptable way" for the laboratory to assess the quality of a method. Further, it places unnecessary burden on manufacturers to make their methods completely interference free, because it absolves the laboratory from testing for interferences (in direct contradiction of CLIA).

- **There are several warnings about using quality control samples in Section 5.2.**

Section 5.2.1

20. I believe the section on use of control materials instead of patient specimens should be omitted. Even though there are several caveats in the present text, there are so many reasons against doing this that it makes no sense to encourage the practice. Laboratories that have no access to patient samples probably have little use for the protocol in the first place. They can analyze their QC results by standard approaches/metrics.

- **The subcommittee (and area committee) have had trouble agreeing on how to use (if at all) quality control materials. The section has been changed to clarify that use of quality control materials is not recommended as an alternative to patient samples. There may be some benefit from analyzing existing quality control data for monitoring total analytical error.**

Section 6

21. Subsection titling is not consistent in Section 6 and could lead to some reader confusion. I recommend reorganizing the topic hierarchy to focus squarely on the proposed analysis approaches as:

    6.2 Total analytical error Estimation
    6.2.1 Introduction (incorporate information from current Sections 6.6 and 6.7 into this section)
    6.2.2 Outliers
    6.2.3 Non-Parametric Analysis

- **The subcommittee believes that the current titling is adequate.**

Section 6.1.1.1

22. In their paper, Bland and Altman do not provide reasons for using the mean instead of the X-values alone (they refer readers to a reference). Further, difference plots are useful only for looking at trends, not for acceptance or rejection of a method. In addition, the Bland and Altman approach is only for nonanalytical methods, such as comparing two different types of blood pressure measurements (at least according to their paper).

- **Interested readers can consult the reference cited by Bland and Altman to see why the mean of the X- and Y-values is recommended if the X-method is not a reference method. Inclusion of this statistical discussion was not thought to be of interest to most users of NCCLS documents. The difference plot has more information than merely a trend. Looking at the plot alone is not suggested as the sole means to accept or reject a method. The use of difference plots is widely accepted for quantitative methods.**

Section 6.1.1.2

23. Expand the text in Section 6.1.1.2 from "…random error increases with concentration," to "…random error changes with concentration." This will extend the description to cover assays like FT4 and FT3 that show increased TAE at lower analyte concentrations.

- **This change has been incorporated.**

Section 6.1.2

24. The mountain plot is a less sensitive and specific method for analyzing errors than traditional methods (i.e., calculating mean and SD for precision; examining slope, intercept, and predicted bias from linear regression analysis). Further, it requires specific software to generate, not available in most laboratories.

- **The mountain plot provides different insights into the data. The use of traditional analysis methods is not precluded. Instructions are provided for performing calculations, which can be performed using spreadsheet software.**

Section 6.2

25. This section assumes that nonparametric methods are superior to parametric ones. They are not. They require many more data points to establish and have much weaker power distributions. They should be used only when it is obvious that the errors one is encountering are due to obviously non-Gaussian distributions. This section advises one to "be aware" that the sample is representative of the population and that there are errors involved; however, it fails to give specifics on these errors and how to detect them. Furthermore, the mountain plot technique is not generally accepted, nor has it undergone scrutiny by disinterested parties.

- **Section 6.2 does not state that nonparametric methods are superior to parametric ones. The mountain plot has been published in a statistical journal (see Reference 8) and a clinical chemistry journal (see Reference 7). It is true that mountain plot analysis is not widely used. Difference plots took some time to gain acceptance and still compete with scatter plots even though difference plots are superior.**

Section 6.4

26. The test described herein is a weak test for observing differences between two data sets.

- **The rationale for using the nonparametric test is described in the first paragraph of the section. Although it is true that parametric tests are more efficient statistically, these differences diminish with increasing sample sizes.**

27. The second sentence in Section 6.4 should read: "It, however, assumes random sampling and the inclusion of pertinent sources of variation in the experiment."

- **This change has been incorporated.**

28. The third sentence in that section should replace the word "minimum" with "minimal."

- **This change has been incorporated.**

Section 6.5

29. Even with moderate sample sizes, the histogram can be quite variable. A technique that provides easier interpretation, particularly in small sample sizes, is the normal probability plot.

- **This change has been incorporated.**

Section 6.6

30. The penultimate paragraph in Section 6.6 should be deleted. If so much data is available that simply calculating the SD is appropriate, then why bother utilizing the methods suggested in References 13 through 18?

- **The direct method of estimating total analytical error (Sections 6.4 and 6.5) does not provide information about the magnitude of the individual error sources. However, this information is provided by simulation. This will be of value to manufacturers and some laboratories. The section mentioned by the commenter cautions that the simulation should be checked by comparing its results with a direct estimate of total analytical error.**

31. An incorrect box title from Figure 1 is used in Section 6.6. The title "Proportional Bias" is a holdover from an earlier version of Figure l (CLN, March 2001); in this document it should read "Linear Bias."

- **This has been changed. "Proportional Bias" was also added as it is still used.**

Section 7.3.2

32. The formula for calculation of percentile is wrong in Sections 7.3.2 and 7.4.2. The denominator should refer to Column E (ranks), not Column D (differences).

- **The subcommittee thanks the commenter for noting these errors. The appropriate changes have been made to the text.**

Section 7.3.5

33. Include the spreadsheet column "Ids" (e.g., A, B, C, . . .) in Tables 1 and 2. This will facilitate going between the formulae (Sections 7.3.2 and 7.4.2) and the examples.

- **This recommendation has been adopted.**

Section 7.4.5

34. Use a consistent number of decimal figures in the data of Table 2, Columns A, B, and C.

- **This recommendation has been adopted.**

Summary of Delegate Comments and Subcommittee Responses

EP21-A: *Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline*

Section 3

1. For the term "accuracy," we recommend addition of text or a footnote that this is called "trueness" in ISO terms.

- **The term "accuracy" is defined correctly in the Definitions section and used throughout EP21 consistent with ISO usage. No footnote is necessary.**

2. For the term "coefficient of variation," we recommend running the two definitions together, i.e., "A measure of relative precision. Most commonly, for a non-negative characteristic, the ration of the standard…" Also, we suggest including in the note that this is most meaningful for normal distributions.

- **Both definitions are accepted ISO/VIM definitions. For better clarity, however, the subcommittee has decided to use only the second definition, which is more appropriate for this guideline.**

Section 4.3

3. In the second to last sentence, eliminate the parenthetical editorial comments or change "they will" to "they can."

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

Section 5.1

4. Specimens containing interfering substances should not be included if the level of interferent exceeds the limits claimed by the manufacturer. Perhaps a note could be added to instruct the user to be sure to consult the package insert for allowable specimens and to consider the limitations provided therein.

- **Although the commenter is correct in principle, it is impractical for a laboratory to know the concentration of every interfering substance listed by the manufacturer. However, a sentence has been added for clarity.**

Section 5.2

5. It is not clear how to work with controls.

- **Controls are treated no differently than patient samples.**

Section 5.2.1

6. The explanation given in this section is not enough to prepare the Mountain plot. If outliers are found and removed, does this mean that the total analytical error interval would change?

- **Please see the examples in Section 7.3, Example 1. LDL Cholesterol. There is no basis for removing outliers, nor is it recommended.**

Section 6.1.1 and Section 6.1.2

7. Some readers may not be familiar with these graphs. Refer the reader to the examples provided in Figures 2, 3, 4, and 5.

- **Readers are referred to examples in Sections 6.1.1, The Difference Plot and 6.1.2, The Mountain Plot.**

Section 6.2

8. Representative specimens must be within the manufacturer's intended use and indications for use claims.

- **While this statement is true, it is unclear how it relates to the estimation of total analytical error or to this section.**

Section 6.6

9. The method to calculate total analytical error is very interesting. However, for analytes with wide ranges (e.g., TSH), it will be necessary to prepare three or four differences charts. It means at least 120 samples.

- **The use of percent differences rather than absolute differences may reduce the number of charts needed.**

Section 7.3.2

10. I believe that there is a typographical error. "Column G is the folded percentile = Column ***F*** (not G as written) for all percentiles…"

- **This typographical correction has been made.**

Section 7.4.2

11. The sentence "Column F contains…which is 100 in this example" should read, "which is 125 in this example."

- **This correction has been made.**

Figure 5

12. Expand the X-axis to include the limit lines for ±4. Also, we recommend that the limit lines be made a distinctly different pattern in all figures to facilitate reading of the graphs.

- **The subcommittee thanks the commenter for this suggestion. The plots depicted in this version of EP21 were generated by software that is currently programmed to scale the axis by the data, not the limits. This recommendation will be considered during the next revision of the document.**

The Quality System Approach

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

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

EP21-A addresses the quality system essentials (QSEs) indicated by an "X." For a description of the other NCCLS documents listed in the grid, please refer to the Related NCCLS Publications section below.

Documents & Records	Organization	Personnel	Equipment	Purchasing & Inventory	Process Control	Information Management	Occurrence Management	Assessment	Process Improvement	Service & Satisfaction	Facilities & Safety
					X GP10						

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

Related NCCLS Publications*

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

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

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

NOTES

NOTES

NOTES

NOTES

Active Membership
(as of 1 April 2003)

Sustaining Members

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

Professional Members

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

Government Members

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

Industry Members

AB Biodisk
Abbott Laboratories
Abbott Laboratories, MediSense Products
Acrometrix Corporation
Alifax S.P.A.
Ammirati Regulatory Consulting
Anaerobe Systems
A/S ROSCO
Asséssor
AstraZeneca Pharmaceuticals
Avant Immunotherapeutics, Inc.
Aventis
Axis-Shield POC AS
Bayer Corporation – Elkhart, IN
Bayer Corporation – Tarrytown, NY
Bayer Corporation – West Haven, CT
Bayer Medical Ltd.
BD
BD Biosciences – San Jose, CA
BD Consumer Products
BD Diagnostic Systems
BD Italia S.P.A.
BD VACUTAINER Systems
Beckman Coulter, Inc.
Beckman Coulter, Inc. Primary Care Diagnostics
Beckman Coulter K.K. (Japan)
Bio-Development SRL
Bio-Inova Life Sciences International
Bio-Inova Life Sciences North America
BioMérieux (NC)
bioMérieux, Inc. (MO)
Biometrology Consultants
Bio-Rad Laboratories, Inc.
Bio-Rad Laboratories, Inc. - France
Blaine Healthcare Associates, Inc.
Bristol-Myers Squibb Company
Canadian External Quality Assessment Laboratory
Chiron Corporation
ChromaVision Medical Systems, Inc.
Clinical Design Group Inc.
Cognigen
Community Medical Center (NJ)
Copan Diagnostics Inc.
Cosmetic Ingredient Review
Cubist Pharmaceuticals
Dade Behring Inc. - Deerfield, IL
Dade Behring Inc. - Glasgow, DE
Dade Behring Inc. - Marburg, Germany
Dade Behring Inc. - Sacramento, CA
Dade Behring Inc. - San Jose, CA
David G. Rhoads Associates, Inc.
Diagnostics Consultancy
Diagnostic Products Corporation
Eiken Chemical Company, Ltd.
Elan Pharmaceuticals
Electa Lab s.r.l.
Enterprise Analysis Corporation
Essential Therapeutics, Inc.
EXPERTech Associates, Inc.
F. Hoffman-La Roche AG
Fort Dodge Animal Health
General Hospital Vienna (Austria)
Gen-Probe
GlaxoSmithKline
Greiner Bio-One Inc.
IGEN Inc.
Immunicon Corporation
Instrumentation Laboratory
International Technidyne Corporation
IntraBiotics Pharmaceuticals, Inc.
I-STAT Corporation
Johnson and Johnson Pharmaceutical Research and Development, L.L.C.
LAB-Interlink, Inc.
Laboratory Specialists, Inc.
Labtest Diagnostica S.A.
LifeScan, Inc. (a Johnson & Johnson Company)
Lilly Research Laboratories
Medical Device Consultants, Inc.
Merck & Company, Inc.
Minigrip/Zip-Pak
mvi Sciences (MA)
Nichols Institute Diagnostics (Div. of Quest Diagnostics, Inc.)
NimbleGen Systems, Inc.
Nissui Pharmaceutical Co., Ltd.
Nippon Becton Dickinson Co., Ltd.
Norfolk Associates, Inc.
Novartis Pharmaceuticals Corporation
Ortho-Clinical Diagnostics, Inc. (Rochester, NY)
Ortho-McNeil Pharmaceutical
Oxoid Inc.
Paratek Pharmaceuticals
Pfizer Inc
Pharmacia Corporation
Philips Medical Systems
Powers Consulting Services
Premier Inc.
Procter & Gamble Pharmaceuticals, Inc.
QSE Consulting
Quintiles, Inc.
Radiometer America, Inc.
Radiometer Medical A/S
Replidyne
Roche Diagnostics GmbH
Roche Diagnostics, Inc.
Roche Laboratories (Div. Hoffmann-La Roche Inc.)
Sarstedt, Inc.
SARL Laboratoire Carron (France)
Schering Corporation
Schleicher & Schuell, Inc.
Second Opinion
Seraphim Life Sciences Consulting LLC
Streck Laboratories, Inc.
Synermed Diagnostic Corp.
Sysmex Corporation (Japan)
Sysmex Corporation (Long Grove, IL)
The Clinical Microbiology Institute
The Toledo Hospital (OH)
Theravance Inc.
Transasia Engineers
Trek Diagnostic Systems, Inc.
Tyco Kendall Healthcare
Versicor, Inc.
Vetoquinol S.A.
Vysis, Inc.
Wyeth Research
Xyletech Systems, Inc.
YD Consultant
YD Diagnostics (Seoul, Korea)

Trade Associations

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

Associate Active Members

31st Medical Group/SGSL (APO, AE)
67th CSH Wuerzburg, GE (NY)
121st General Hospital (CA)
Academisch Ziekenhuis-VUB (Belgium)
Acadiana Medical Laboratories, LTD (LA)
Advocate Healthcare Lutheran General (IL)
Akershus Central Hospital and AFA (Norway)
Albemarle Hospital (NC)
Allegheny General Hospital (PA)
Allina Health System (MN)
Alton Ochsner Medical Foundation (LA)
Anne Arundel Medical Center (MD)
Antwerp University Hospital (Belgium)
Arkansas Department of Health
ARUP at University Hospital (UT)
Armed Forces Research Institute of Medical Science (APO, AP)
Associated Regional & University Pathologists (UT)
Atlantic Health System (NJ)
Aurora Consolidated Laboratories (WI)
AZ Sint-Jan (Belgium)
Azienda Ospedale Di Lecco (Italy)
Baxter Regional Medical Center (AR)
Bay Medical Center (MI)
Baystate Medical Center (MA)
Bbaguas Duzen Laboratories (Turkey)
BC Biomedical Laboratories (Surrey, BC, Canada)
Bermuda Hospitals Board
Bo Ali Hospital (Iran)
Brooks Air Force Base (TX)
Broward General Medical Center (FL)
Cadham Provincial Laboratory
Calgary Laboratory Services
Cape Breton Healthcare Complex (Nova Scotia, Canada)
Carilion Consolidated Laboratory (VA)
Cathay General Hospital (Taiwan)
Cavan General Hospital (Ireland)
Central Peninsula General Hospital (AK)
Central Texas Veterans Health Care System
Centro Diagnostico Italiano (Milano, Italy)
Champlain Valley Physicians Hospital (NY)
Chang Gung Memorial Hospital (Taiwan)
Changi General Hospital (Singapore)
The Charlotte Hungerford Hospital (CT)
Children's Hospital (LA)
Children's Hospital (NE)
Children's Hospital & Clinics (MN)
Children's Hospital Medical Center (Akron, OH)
Children's Hospital of Philadelphia (PA)
Children's Medical Center of Dallas (TX)
CHR St. Joseph Warquignies (Belgium)
Clarian Health–Methodist Hospital (IN)
CLSI Laboratories (PA)
Commonwealth of Kentucky
Community Hospital of Lancaster (PA)
CompuNet Clinical Laboratories (OH)
Cook County Hospital (IL)
Cook Children's Medical Center (TX)
Covance Central Laboratory Services (IN)
Danish Veterinary Laboratory (Denmark)
Danville Regional Medical Center (VA)
Dean Medical Center (WI)
Department of Health & Community Services (New Brunswick, Canada)
DesPeres Hospital (MO)
Detroit Health Department (MI)
Diagnosticos da América S/A (Brazil)
Dianon Systems, Inc.
Dr. Everett Chalmers Hospital (New Brunswick, Canada)
Doctors Hospital (Bahamas)
Duke University Medical Center (NC)
Dwight David Eisenhower Army Med. Ctr. (GA)
E.A. Conway Medical Center (LA)
East Side Clinical Laboratory (RI)
Eastern Health (Vic., Australia)
Elyria Memorial Hospital (OH)
Emory University Hospital (GA)
Fairview-University Medical Center (MN)

Federal Medical Center (MN)
Florida Hospital East Orlando
Focus Technologies (CA)
Focus Technologies (VA)
Foothills Hospital (Calgary, AB, Canada)
Fresenius Medical Care/Spectra East (NJ)
Fresno Community Hospital and Medical Center
Frye Regional Medical Center (NC)
Gambro BCT (CO)
Gamma Dynacare Medical Laboratories (ON, Canada)
Geisinger Medical Center (PA)
Grady Memorial Hospital (GA)
Guthrie Clinic Laboratories (PA)
Hahnemann University Hospital (PA)
Harris Methodist Fort Worth (TX)
Hartford Hospital (CT)
Health Network Lab (PA)
Health Partners Laboratories (VA)
Highlands Regional Medical Center (FL)
Hoag Memorial Hospital Presbyterian (CA)
Holy Cross Hospital (MD)
Holmes Regional Medical Center (FL)
Holzer Medical Center (OH)
Hôpital du Sacré – Coeur de Montreal (Montreal, Quebec, Canada)
Hôpital Maisonneuve – Rosemont (Montreal, Canada)
Hôpital Saint-LUC (Montreal, Quebec, Canada)
Hospital for Sick Children (Toronto, ON, Canada)
Hospital Sousa Martins (Portugal)
Hotel Dieu Hospital (Windsor, ON, Canada)
Huddinge University Hospital (Sweden)
Hurley Medical Center (MI)
Indiana University
Innova Fairfax Hospital (VA)
Institute of Medical and Veterinary Science (Australia)
International Health Management Associates, Inc. (IL)
Jackson Memorial Hospital (FL)
Jersey Shore Medical Center (NJ)
John F. Kennedy Medical Center (NJ)
John Peter Smith Hospital (TX)
Kadlec Medical Center (WA)
Kaiser Permanente Medical Care (CA)
Kaiser Permanente (MD)
Kantonsspital (Switzerland)
Kenora-Rainy River Regional Laboratory Program (Ontario, Canada)
Kimball Medical Center (NJ)
King's Daughter Medical Center (KY)
Klinični Center (Slovenia)
LabCorp (NC)
Laboratoire de Santé Publique du Quebec (Canada)
Laboratorio Dr. Echevarne (Spain)
Laboratório Fleury S/C Ltda. (Brazil)
Laboratory Corporation of America (NJ)
LAC and USC Healthcare Network (CA)
Lakeland Regional Medical Center (FL)
Lancaster General Hospital (PA)
Langley Air Force Base (VA)
LeBonheur Children's Medical Center (TN)
L'Hotel-Dieu de Quebec (Canada)
Libero Instituto Univ. Campus BioMedico (Italy)
Lourdes Health System (NJ)
Louisiana State University Medical Center
Maccabi Medical Care and Health Fund (Israel)
Malcolm Grow USAF Medical Center (MD)
Martin Luther King/Drew Medical Center (CA)
Massachusetts General Hospital (Microbiology Laboratory)
MDS Metro Laboratory Services (Burnaby, BC, Canada)
Medical College of Virginia Hospital
Medicare/Medicaid Certification, State of North Carolina
Memorial Medical Center (IL)
Memorial Medical Center (LA) Jefferson Davis Hwy
Memorial Medical Center (LA) Napoleon Avenue
Mercy Medical Center (IA)
Methodist Hospital (TX)
MetroHealth Medical Center (OH)
Michigan Department of Community Health
Mississippi Baptist Medical Center
Monte Tabor – Centro Italo - Brazileiro de Promocao (Brazil)
Montreal Children's Hospital (Canada)
Montreal General Hospital (Canada)
National Institutes of Health (MD)
National University Hospital (Singapore)
Naval Hospital – Corpus Christi (TX)
Nebraska Health System
New Britain General Hospital (CT)
New England Fertility Institute (CT)
New England Medical Center (MA)
New Mexico VA Health Care System
New York University Medical Center
NorDx (ME)
North Carolina State Laboratory of Public Health
North Shore – Long Island Jewish Health System Laboratories (NY)
North Shore University Hospital (NY)
Northwestern Memorial Hospital (IL)
O.L. Vrouwziekenhuis (Belgium)
Ordre professionnel des technologists médicaux du Québec
Ospedali Riuniti (Italy)
The Ottawa Hospital (Ottawa, ON, Canada)
OU Medical Center (OK)
Pathology Associates Medical Laboratories (WA)
The Permanente Medical Group (CA)
Piedmont Hospital (GA)
Pocono Medical Center (PA)
Presbyterian Hospital of Dallas (TX)
Providence Health Care
Queen Elizabeth Hospital (Prince Edward Island, Canada)
Queensland Health Pathology Services (Australia)
Quest Diagnostics Incorporated (CA)
Quintiles Laboratories, Ltd. (GA)
Regions Hospital
Reid Hospital & Health Care Services (IN)
Research Medical Center (MO)
Rex Healthcare (NC)
Rhode Island Department of Health Laboratories
Riverside Medical Center (IL)
Riyadh Armed Forces Hospital (Saudi Arabia)
Robert Wood Johnson University Hospital (NJ)
Royal Columbian Hospital (New Westminster, BC, Canada)
Saad Specialist Hospital (Saudi Arabia)
Sacred Heart Hospital (MD)
Saint Mary's Regional Medical Center (NV)
St. Alexius Medical Center (ND)
St. Anthony Hospital (CO)
St. Anthony's Hospital (FL)
St. Barnabas Medical Center (NJ)
St-Eustache Hospital (Quebec, Canada)
St. Francis Medical Ctr. (CA)
St. John Hospital and Medical Center (MI)
St. John Regional Hospital (St. John, NB, Canada)
St. John's Hospital & Health Center (CA)
St. Joseph Hospital (NE)
St. Joseph's Hospital – Marshfield Clinic (WI)
St. Joseph's Medical Center (NY)
St. Jude Children's Research Hospital (TN)
St. Luke's Regional Medical Center (IA)
St. Mary of the Plains Hospital (TX)
St. Mary's Hospital & Medical Center (CO)
St. Michael's Hospital (Toronto, ON, Canada)
St. Vincent Medical Center (CA)
Ste. Justine Hospital (Montreal, PQ, Canada)
Salina Regional Health Center (KS)
San Francisco General Hospital (CA)
Santa Clara Valley Medical Center (CA)
Sentara Williamsburg Community Hospital (VA)
Seoul Nat'l University Hospital (Korea)
South Bend Medical Foundation (IN)
So. California Permanente Medical Group
Southwest Texas Methodist Hospital (TX)
South Western Area Pathology Service (Australia)
Southern Maine Medical Center
Spartanburg Regional Medical Center (SC)
Specialty Laboratories, Inc. (CA)
Stanford Hospital and Clinics (CA)
State of Washington Department of Health
Stony Brook University Hospital (NY)
Stormont-Vail Regional Medical Center (KS)
Sun Health-Boswell Hospital (AZ)
Sunnybrook Health Science Center (ON, Canada)
Swedish Medical Center – Providence Campus (WA)
Temple University Hospital (PA)
Tenet Odessa Regional Hospital (TX)
The Toledo Hospital (OH)
Touro Infirmary (LA)
Trident Regional Medical Center (SC)
Tripler Army Medical Center (HI)
Truman Medical Center (MO)
Tuenmun Hospital (Hong Kong)
UCLA Medical Center (CA)
UCSF Medical Center (CA)
UNC Hospitals (NC)
University College Hospital (Galway, Ireland)
University Hospital (Gent) (Belgium)
University Hospitals of Cleveland (OH)
University of Alabama-Birmingham Hospital
University of Alberta Hospitals (Canada)
University of Colorado Health Science Center
University of Chicago Hospitals (IL)
University of Illinois Medical Center
University of the Ryukyus (Japan)
University of Virginia Medical Center
UZ-KUL Medical Center (Belgium)
VA (Hines) Medical Center
VA (Kansas City) Medical Center (MO)
VA (San Diego) Medical Center (CA)
VA (Tuskegee) Medical Center (AL)
Vejle Hospital (Denmark)
Virginia Department of Health
ViroMed Laboratories (MN)
Warren Hospital (NJ)
Washington Adventist Hospital (MD)
Washoe Medical Center Laboratory (NV)
Waterford Regional Hospital (Ireland)
Wellstar Health Systems (GA)
West Jefferson Medical Center (LA)
West Shore Medical Center (MI)
Wilford Hall Medical Center (TX)
William Beaumont Army Medical Center (TX)
William Beaumont Hospital (MI)
William Osler Health Centre (Brampton, ON, Canada)
Winn Army Community Hospital (GA)
Winnipeg Regional Health Authority (Winnipeg, Canada)
Wishard Memorial Hospital (IN)
Yonsei University College of Medicine (Korea)
York Hospital (PA)

NCCLS ▼ 940 West Valley Road ▼ Suite 1400 ▼ Wayne, PA 19087 ▼ USA ▼ PHONE 610.688.0100
FAX 610.688.0700 ▼ E-MAIL: exoffice@nccls.org ▼ WEBSITE: www.nccls.org ▼ ISBN 1-56238-502-X