

EP9-A2
ISBN 1-56238-472-4
ISSN 0273-3099

Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition

Volume 22 Number 19

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



Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition

Abstract

NCCLS document EP9-A2—*Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition*, is written for laboratorians as well as manufacturers. It describes procedures for determining the relative bias between two methods, and it identifies factors to be considered when designing and analyzing a method-comparison experiment using split patient samples. For carrying out method-comparison evaluations, an overview of the experiment, sample data recording and calculation sheets, and an overview flowchart and a detailed flowchart for preliminary data examination are included. As an additional aid, a sample scatter plot and bias plot are introduced for those who are unfamiliar with these procedures. The final section contains recommendations for manufacturers' evaluation of bias and statement format for bias claims.

NCCLS. *Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition*. NCCLS document EP9-A2 (ISBN 1-56238-472-4). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.

THE NCCLS consensus process, which is the mechanism for moving a document through two or more levels of review by the healthcare community, is an ongoing process. Users should expect revised editions of any given document. Because rapid changes in technology may affect the procedures, methods, and protocols in a standard or guideline, users should replace outdated editions with the current editions of NCCLS documents. Current editions are listed in the *NCCLS Catalog*, which is distributed to member organizations, and to nonmembers on request. If your organization is not a member and would like to become one, and to request a copy of the *NCCLS Catalog*, contact the NCCLS Executive Offices. Telephone: 610.688.0100; Fax: 610.688.0700; E-Mail: exoffice@nccls.org; Website: www.nccls.org

Contents

Abstract	i
Committee Membership	v
Active Membership	vii
Foreword	xv
The Quality System Approach	xvi
1 Introduction and Scope	1
1.1 Overview of the General Comparison Experiment	1
1.2 Symbols Used in the Text	2
1.3 Definitions	3
2 Device-Familiarization Period	4
3 Comparison of Methods Experiment	4
3.1 Test Samples	4
3.2 Comparative Method	4
3.3 Range of Measurement	5
3.4 Number of Samples	5
3.5 Sample Sequence	6
3.6 Time and Duration	6
3.7 Inspection of Data During Collection	6
3.8 Quality Control	6
3.9 Documentation of Rejected Data	7
4 Preliminary Data Examination	7
4.1 Outlier Tests on Within-Method Duplicates	11
4.2 Plotting the Data	12
4.3 Visual Check for Linear Relationship	12
4.4 Visual Check for Between-Method Outliers	12
4.5 Test for Adequate Range of X	13
5 Linear Regression	14
5.1 Computations	14
5.2 Visual Check for Constant Scatter	15
6 Computing Predicted Bias and Its Confidence Interval	16
6.1 Linear Regression Procedure (When Data Pass Adequate Range and Uniform Scatter Checks)	16
6.2 Computing Average Bias Using Partitioned Individual Differences When Data Fail Adequate Range Check (Partitioned Biases Procedure)	17
6.3 Computing Predicted Bias Using Partitioned Residuals When Data Have Nonconstant (Variable) Precision (Partitioned Residuals Procedure)	18
7 Interpreting Results and Comparing to Internal Performance Criteria	18
8 Manufacturer Modifications	19

Contents (Continued)

8.1 Experimental Design19

8.2 Data Analysis.....19

8.3 Statement of Bias Performance Claims19

References.....23

Appendix A. Sample Data Recording Sheet.....24

Appendix B. Scatter Plots Derived from Example26

Appendix C. Calculation Example.....30

Appendix D. Calculation of Deming Slope35

Summary of Comments and Working Group Responses36

Summary of Delegate Comments and Committee Responses.....50

Related NCCLS Publications.....53

Foreword

The current literature contains many examples of user and manufacturer product evaluations, with many different experimental and statistical procedures¹ for comparing two methods that measure the same analyte. This methodologic variety has caused confusion, and users have reported that comparisons often lack sufficient data and description to be reproducible.

There has also been an increasing awareness that the scope of evaluation procedures appropriate for manufacturers of diagnostic devices is not always appropriate for their users. The manufacturer is concerned with establishing valid and achievable performance claims for bias when compared with a generally accepted standard or reference method. The user might wish to compare a candidate method with a different one than the manufacturer used in establishing the bias claims. The scope of the experimental and data-handling procedures for these two purposes can often differ.

Therefore, in preparing this document, the working group drew on the experience of users and representatives of industry, statisticians, and laboratory and medical personnel. Because of the many *in vitro* diagnostic methods and kits now available, the working group realizes that a single experimental design is not appropriate for all types of user and manufacturer method comparisons. Therefore, this guideline was developed primarily to give conceptual help in structuring an experiment for comparing two methods. To illustrate representative duration, procedures, materials, methods of quality control, statistical data handling, and interpretation of results, an example experiment is presented.

Throughout the development of this protocol, the working group had to decide which procedural and statistical methods to recommend in the example experiment. To respond to the needs of laboratorians and manufacturers, the working group combined input from users of analytical methods, manufacturers of these methods, and representatives of regulatory agencies. The working group also included the recommendations necessary for a scientifically valid comparison. Compromises were necessary to accommodate both the simplicity of operation protocol and the complexity of design and statistical calculations necessary for valid conclusions. This document is adaptable within a wide range of analytes and device complexity.

The focus of this document is the independent establishment of bias performance characteristics. If appropriate, the user is then free to compare these performance estimates with either the manufacturer's labeled claims or the user's own internal criteria.

The working group believes that standard experimental and statistical procedures in user method comparisons will make such evaluations more reproducible and reflective of actual performance, and the statements of evaluation results considerably more reliable. Also, the misuse and misinterpretation of statistical methods, such as regression and correlation, involved in comparing *in vitro* diagnostic devices can seriously impair the usefulness of such evaluations. Therefore, this document is intended to promote the effective use of statistical analysis and data reporting.

Manufacturers of laboratory devices are encouraged to use this guideline to establish and standardize their bias performance claims. Many different forms have been used for such claims, and they have not always been sufficiently specific to allow user verification.

Key Words

Bias, evaluation protocol, experimental design, linear regression, method comparison, quality control, residuals

The Quality System Approach

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

QSEs

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

EP9-A2 Addresses the following Quality System Essentials (QSEs)

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

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

Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition

1 Introduction and Scope

This document provides both users and manufacturers of clinical laboratory devices with guidance for designing an experiment to evaluate the bias between two methods that measure the same analyte. Ideally, a test (or candidate) method should be compared with a reference method. For users, the comparative method is often the current routine method, however, and the purpose of the evaluation is to determine if the two methods yield equivalent results within the statistical power of the experiment. In this case, determining whether the test method is a suitable replacement for a current method is the primary concern.

This guideline allows the estimation of the bias (expected difference) between two methods at various concentrations. If the comparative method is the same one used by the manufacturer in the statement of claims, it is possible to compare statistically the experimental results to the manufacturer's claims to verify acceptable performance.

1.1 Overview of the General Comparison Experiment

Evaluating an analytical method requires the following:

- Sufficient time for the operators to become familiar with the device's operation and maintenance procedures.
- Sufficient time for the operators to become familiar with the evaluation protocol.
- Assurance that both the test and the comparative methods are in proper quality control throughout the evaluation period.
- Sufficient data to ensure representative results for both the test and the comparative methods. (What constitutes sufficient data will depend on the precision and interference effects of the two methods, the amount of bias between the two methods, the range of sample analyte values available, and the medical requirements of the test.)

During the device familiarization period, the operators of the test and comparative methods must become familiar with all aspects of set-up, operation, maintenance, trouble-shooting, and quality control of both methods. This period can precede other parts of the evaluation process or coincide with the manufacturer's training period. Run routine laboratory quality control procedures on both methods.

After the familiarization period, the method-comparison experiment can begin. The working group recommends that at least 40 patient samples be analyzed over at least 5 operating days. The reliability and effectiveness of the experiment increase by analyzing more samples over more time, while following the manufacturer's recommendations for calibration.

Analyze each patient sample in duplicate using both the test method and the comparative method. Analyze the duplicates for each method within the same run for that method. Whenever possible, at least 50% of the samples run should be outside the laboratory's reference interval.

When the experiment is completed, record the data in a logical manner (such as that which is suggested in the Appendix). Plot the data and assess the diagram visually and statistically for relative linearity,

adequate range, and uniform scatter. Based on the results of the data examination, use either simple linear regression or alternative procedures to estimate expected (average) bias and the confidence interval for expected bias at any desired medical decision level. Then, these estimates can be compared with claims or internal criteria to judge the acceptability of the method.

1.2 Symbols Used in the Text

The following symbols are used in this document:

X	comparative method
Y	test method
DX_i or DY_i	absolute value of the difference between duplicates for method X or Y
i	sample number
N	total number of samples
1,2 or j	duplicate number or replicate number (as a subscript)
\overline{DX} or \overline{DY}	mean absolute difference of method
DX'_i or DY'_i	normalized (relative) absolute difference of method
$\overline{DX'}$ or $\overline{DY'}$	normalized (relative) mean absolute difference of method
E_{ij}	absolute difference between methods
\overline{E}	mean absolute difference between methods
E'_{ij}	relative absolute difference between methods
$\overline{E'}$	relative mean absolute difference between methods
TL_E	test limit
r	correlation coefficient
x	observation from comparative method
y	observation from test method
x_{ij} or y_{ij}	observation (x or y) from run i, replicate j
\bar{x} or \bar{y}	average of x or y
b	slope

Related NCCLS Publications*

- EP5-A** **Evaluation of Precision Performance of Clinical Chemistry Devices; Approved Guideline (1999).** This document provides guidance for designing an experiment to evaluate the precision performance of clinical chemistry devices; recommendations on comparing the resulting precision estimates with manufacturer's precision performance claims and determining when such comparisons are valid, as well as manufacturer's guidelines for establishing claims.
- EP6-P2** **Evaluation of the Linearity of Quantitative Analytical Methods; Proposed Guideline—Second Edition (2001).** This document provides guidelines for characterization the linearity of a method during a method evaluation; for checking linearity as part of routine quality assurance; and for determining and stating a manufacturer's claim for linear range.
- EP7-P** **Interference Testing in Clinical Chemistry; Proposed Guideline (1986).** This document provides background information and procedures for characterizing the effects of interfering substances on test results.
- NRSCL8-A** **Terminology and Definitions For Use in NCCLS Documents; Approved Standard (1998).** This document provides standard definitions for use in NCCLS standards and guidelines, and for submitting candidate reference methods and materials to the National Reference System for the Clinical Laboratory (NRSCL).

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