Chapter 9: Conclusion

Techniques for comparing IVD measurement procedures and quantifying differences between them are described in this guideline. It starts with preparing for and performing studies for collecting the needed data. Examples of difference plots and scatterplots are provided as ways to visualize this comparison data. Measurements of bias based on difference plots can be used by laboratories for verification studies but regression analyses based on scatter plots are necessary for quantification of differences by IVD manufacturers. Such regression analyses are extensively covered including which techniques to use depending upon data characteristics revealed by data visualization. The fit of the selected regression model can then be used to determine bias at result values corresponding to medical decision thresholds. Various situations are described for using these techniques within a measurement procedure. Finally, once differences have been determined, suggestions are made for stating the outcome of the measurement procedure comparison.

Chapter 10: Supplemental Information

This chapter includes:

- References
- Appendixes
- The Quality Management System Approach
- Related CLSI Reference Materials

References

1 CLSI. *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline—Second Edition.* CLSI document EP12-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.

2 CLSI. *Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline—Third Edition.* CLSI document EP05-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

3 CLSI. *User Verification of Precision and Estimation of Bias; Approved Guideline—Third Edition.* CLSI document EP15-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

4 CLSI. *Evaluation of Total Analytical Error for Quantitative Medical Laboratory Measurement Procedures.* 2nd ed. CLSI document EP21. Wayne, PA: Clinical and Laboratory Standards Institute; 2016.

5 CLSI. *Interference Testing in Clinical Chemistry.* 3rd ed. CLSI guideline EP07. Wayne, PA: Clinical and Laboratory Standards Institute; 2018.

6 Miller JM, Astles R, Baszler T, et al.; Biosafety Blue Ribbon Panel, Centers for Disease Control and Prevention (CDC). Guidelines for safe work practices in human and animal medical diagnostic laboratories. *MMWR Suppl.* 2012;61(1):1-102.

7 CLSI. *Protection of Laboratory Workers From Occupationally Acquired Infections; Approved Guideline—Fourth Edition.* CLSI document M29-A4. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

8 Bureau International des Poids et Mesures (BIPM). *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms* (VIM, 3rd edition, JCGM 200:2012). http://www.bipm.org/en/publications/guides/vim.html. Accessed March 28, 2018.

9 ISO. *Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions.* ISO 5725-1. Geneva, Switzerland: International Organization for Standardization; 1994.

10 Mandel J. *The Statistical Analysis of Experimental Data.* Mineola, New York: Dover Publications; 1984:282-292.

11 WHO. Expert Committee on Biological Standardization. *Glossary of Terms for Biological Substances Used for Texts of the Requirements.* WHO unpublished document BS/95.1793. Geneva, Switzerland: World Health Organization; 1991.

12 Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *J Clin Chem Clin Biochem.* 1983;21(11):709-720.

13 Passing H, Bablok W. Comparison of several regression procedures for method comparison studies and determination of sample sizes. Application of linear regression procedures for method comparison studies in clinical chemistry, Part II. *J Clin Chem Clin Biochem.* 1984;22(6):431-445.

14 ISO. *Accuracy (trueness and precision) of measurement methods and results – Part 3: Intermediate measures of the precision of a standard measurement method.* ISO 5725-3. Geneva, Switzerland: International Organization for Standardization; 1994.

15 ISO. *Medical laboratories – Requirements for quality and competence.* ISO 15189. Geneva, Switzerland: International Organization for Standardization; 2012.

16 CLSI. *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition.* CLSI document EP28-A3c. Wayne, PA: Clinical and Laboratory Standards Institute; 2010.

17 CLSI. *Evaluation of Commutability of Processed Samples; Approved Guideline—Third Edition.* CLSI document EP14-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

18 Carroll RJ, Ruppert D. The use and misuse of orthogonal regression in linear errors-in-variables models. *Am Stat.* 1996; 50(1):1-6.

19 CLSI. *Quality Assurance for Design Control and Implementation of Immunohistochemistry Assays; Approved Guideline—Second Edition.* CLSI document I/LA28-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2011.

20 Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. *Clin Chem.* 2011;57(1):76-83.

21 Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat.* 2007;17(4):571-582.

22 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-310.

[23] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8(2):135-160.

[24] Krouwer JS. Why Bland-Altman plots should use X, not (Y+X)/2 when X is a reference method. *Stat Med.* 2008;27(5):778-780.

[25] Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet.* 1995;346(8982):1085-1087.

[26] Carstensen B. *Comparing Clinical Measurement Methods: A Practical Guide.* Chichester, United Kingdom: John Wiley & Sons Ltd; 2010.

[27] Hawkins DM. Diagnostics for conformity of paired quantitative measurements. *Stat Med.* 2002;21(13):1913-1935.

[28] Keller T, Butz H, Lein M, et al. Discordance analysis characteristics as a new method to compare the diagnostic accuracy of tests: example of complexed versus total prostate-specific antigen. *Clin Chem.* 2005;51(3):532-539.

[29] CLSI. *Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline.* CLSI document EP06-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2003.

[30] Linnet K. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clin Chem.* 1998;44(5):1024-1031.

[31] CLSI. *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline—Second Edition.* CLSI document EP17-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.

[32] Martin RF. General Deming regression for estimating systematic bias and its confidence interval in method comparison studies. *Clin Chem.* 2000;46(1):100-104.

[33] Davison AC, Hinkley DV. *Bootstrap Methods and their Application.* Cambridge, UK: Cambridge University Press; 1997.

[34] CLSI. *Clinical Evaluation of Immunoassays; Approved Guideline—Second Edition.* CLSI document I/LA21-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.

[35] CLSI. *Verification of Comparability of Patient Results Within One Health Care System; Approved Guideline (Interim Revision).* CLSI document EP31-A-IR. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.

Appendix A. Deming Regression

Abbreviation for Appendix A

SD standard deviation

With measurement errors in both measurement procedures being compared, the linear model describing the relationship between the measurement procedure results *X* and *Y* can be expressed as:

$$Y = a + b(X + \varepsilon_X) + \varepsilon_Y, \quad \text{(A1)}$$

in which:

a, b = intercept and slope of the linear model, and
ε_X, ε_Y = random errors in the *X* and *Y* measurement procedures.

Equation (A1) parameters (a, b) can be estimated with data using regular Deming regression under the following assumptions: the random errors ε_X, ε_Y are independent (across the measurement procedures, specimens, and replicates) and normally distributed with zero averages and constant, measurand-level–independent SDs, $\sigma(\varepsilon_X)$, $\sigma(\varepsilon_Y)$.

The SDs, $\sigma(\varepsilon_X)$, $\sigma(\varepsilon_Y)$, of the random errors are practically constant for measurement procedures with small analytical measuring intervals of the measurand, such as electrolytes. In other cases, the SDs of random measurement errors are often approximately proportional to the measurand level over a large proportion of the measuring interval. In such cases, constant coefficient of variation Deming analysis is more appropriate, as described in Appendix B.

The information on the SDs of the random errors of measurements' approximate constancy or proportionality to the measurand level is often available from the manufacturers' specifications. When such information is not available, it can be obtained by calculating the SDs of the replicated results of measurements for the tested samples, plotting those vs respective replicate averages, and visually examining the graph. The SDs, $\hat{\sigma}(\varepsilon_{X_i})$, $\hat{\sigma}(\varepsilon_{Y_i})$, of the replicate measurements are calculated for each sample using the following equations (assuming the same number of replicates, *R*, for each measurement procedure, *X* and *Y*, and each of the *N* samples tested):

$$\hat{\sigma}(\varepsilon_{X_i}) = \sqrt{\frac{1}{R-1}\sum_{j=1}^{R}(X_{ij} - \bar{X}_i)^2}, \text{ and} \quad \text{(A2)}$$

$$\hat{\sigma}(\varepsilon_{Y_i}) = \sqrt{\frac{1}{R-1}\sum_{k=1}^{R}(Y_{ik} - \bar{Y}_i)^2}, \quad \text{(A3)}$$

in which:

i = sample number; $i = 1, 2, \ldots, N$, and
j, k = replicate number; $j, k = 1, 2, \ldots, R$.

Appendix A. (Continued)

In the calculations below, $\bar{X}_i$,$\bar{Y}_i$ are the replicate averages when the SDs of the random measurement errors are approximately constant, and they are the averages of the logarithms of replicate measurement results when the SDs are approximately proportional to the measurand level. The replicate averages for these measurement results obtained with the *i*th sample ($i = 1, 2, \ldots, N$) are calculated as:

$$\bar{X}_i = \frac{1}{R}\sum_{j=1}^{R} X_{ij} \tag{A4}$$

$$\bar{Y}_i = \frac{1}{R}\sum_{k=1}^{R} Y_{ik} \tag{A5}$$

Equation (A1) can be rewritten for the replicate averages as:

$$\bar{Y}_i = a + b(\bar{X}_i + \varepsilon_X) + \varepsilon_Y, \tag{A6}$$

in which $\varepsilon_{\bar{X}}$, $\varepsilon_{\bar{Y}}$ are random errors of the replicate averages.

Regular Deming regression provides unbiased minimum variance estimates of the equation A6 parameters (a, b)[1] with modified notation; equation for b assumes positive $\hat{\sigma}_{\bar{X}\bar{Y}}$, which is the case with medical laboratory measurement procedures:

$$b = \frac{\hat{\sigma}_{\bar{Y}}^2 - \hat{\lambda}\hat{\sigma}_{\bar{X}}^2 + \sqrt{\left(\hat{\sigma}_{\bar{Y}}^2 - \hat{\lambda}\hat{\sigma}_{\bar{X}}^2\right)^2 + 4\hat{\lambda}\hat{\sigma}_{\bar{X}\bar{Y}}^2}}{2\hat{\sigma}_{\bar{X}\bar{Y}}} \tag{A7}$$

$$a = \bar{\bar{Y}} - b\bar{\bar{X}} \tag{A8}$$

The parameters used in equations (A7) and (A8) are calculated using formulas (A9) to (A14).

$$\hat{\sigma}_{\bar{X}}^2 = \frac{1}{N}\sum_{i=1}^{N}(\bar{X}_i - \bar{\bar{X}})^2 \tag{A9}$$

$$\hat{\sigma}_{\bar{Y}}^2 = \frac{1}{N}\sum_{i=1}^{N}(\bar{Y}_i - \bar{\bar{Y}})^2 \tag{A10}$$

$$\hat{\sigma}_{\bar{X}\bar{Y}} = \frac{1}{N}\sum_{i=1}^{N}(\bar{X}_i - \bar{\bar{X}})(\bar{Y}_i - \bar{\bar{Y}}) \tag{A11}$$

Appendix A. (Continued)

$$\bar{\bar{X}} = \frac{1}{N}\sum_{i=1}^{N}\bar{X}_i \quad \text{(A12)}$$

$$\bar{\bar{Y}} = \frac{1}{N}\sum_{i=1}^{N}\bar{Y}_i \quad \text{(A13)}$$

$$\hat{\lambda} = \frac{\hat{\sigma}^2(\varepsilon_{\bar{Y}})}{\hat{\sigma}^2(\varepsilon_{\bar{X}})} \quad \text{(A14)}$$

in which:

N	=	number of samples used for fitting model (A3).
$\bar{\bar{X}}, \bar{\bar{Y}}$	=	averages across measurement results obtained with X and Y measurement procedures with samples (grand averages).
$\hat{\sigma}^2_{\bar{X}}, \hat{\sigma}^2_{\bar{Y}}, \hat{\sigma}_{\bar{X}\bar{Y}}$	=	average squares and average cross-product of the deviations of the replicate averages of results of measurement obtained with the X and Y measurement procedures from the respective grand averages.
$\hat{\lambda}$	=	ratio of the variances of random errors of the two measurement procedures (within-run or repeatability when data are collected in a single run).

The constant, measurand-level–independent, random error variance estimates, $\hat{\sigma}^2(\varepsilon_X)$, $\hat{\sigma}^2(\varepsilon_Y)$, are calculated as follows[2] (the equations are modified for the same numbers of replicates, R, for both measurement procedures, X and Y, and each of N specimens):

$$\hat{\sigma}^2(\varepsilon_X) = \frac{1}{N(R-1)}\sum_{i=1}^{N}\sum_{j=1}^{R}(X_{ij} - \bar{X}_i)^2 \quad \text{(A15)}$$

$$\hat{\sigma}^2(\varepsilon_Y) = \frac{1}{N(R-1)}\sum_{i=1}^{N}\sum_{k=1}^{R}(Y_{ik} - \bar{Y}_i)^2 \quad \text{(A16)}$$

Each of the above variances has $N(R-1)$ degrees of freedom.

The variances of the averages of R replicate results of measurement are R times smaller than the variances of the individual results given in equations (A15) and (A16):

Appendix A. (Continued)

$$\hat{\sigma}^2(\varepsilon_{\bar{X}}) = \frac{1}{NR(R-1)}\sum_{i=1}^{N}\sum_{j=1}^{R}(X_{ij} - \bar{X}_i)^2 \quad \text{(A17)}$$

$$\hat{\sigma}^2(\varepsilon_{\bar{Y}}) = \frac{1}{NR(R-1)}\sum_{i=1}^{N}\sum_{k=1}^{R}(Y_{ik} - \bar{Y}_i)^2 \quad \text{(A18)}$$

The equations for the estimates of the variances of the intercept, σ_a^2, and slope, σ_b^2, and their covariance, σ_{ab}, in Deming regression (large sample size approximation) are as follows (modified from Miller)[1]:

$$\hat{\sigma}_a^2 = \frac{1}{N}\left[\hat{\sigma}_{\bar{Y}}^2 - 2b\hat{\sigma}_{\bar{X}\bar{Y}} + b^2\hat{\sigma}_{\bar{X}}^2 + \frac{\bar{\bar{X}}^2 b^2}{\sigma_{\bar{X}\bar{Y}}^2}(\hat{\sigma}_{\bar{X}}^2\hat{\sigma}_{\bar{Y}}^2 - \hat{\sigma}_{\bar{X}\bar{Y}}^2)\right] \quad \text{(A19)}$$

$$\hat{\sigma}_b^2 = \frac{b^2}{N\hat{\sigma}_{\bar{X}\bar{Y}}^2}(\hat{\sigma}_{\bar{X}}^2\hat{\sigma}_{\bar{Y}}^2 - \hat{\sigma}_{\bar{X}\bar{Y}}^2) \quad \text{(A20)}$$

$$\hat{\sigma}_{ab} = -\frac{\bar{\bar{X}}b^2}{N\hat{\sigma}_{\bar{X}\bar{Y}}^2}(\hat{\sigma}_{\bar{X}}^2\hat{\sigma}_{\bar{Y}}^2 - \hat{\sigma}_{\bar{X}\bar{Y}}^2) \quad \text{(A21)}$$

Assuming the Y_is follow a normal (gaussian) distribution, the above variances have $N-2$ degrees of freedom, and the $100(1-\gamma)\%$ confidence intervals for the slope and intercept are:

$a \pm t(N-2,1-\frac{\gamma}{2})\hat{\sigma}_a$, and (A22)

$b \pm t(N-2,1-\frac{\gamma}{2})\hat{\sigma}_b$, (A23)

in which σ_a, σ_b = SDs of the intercept and slope estimates found as square roots of the respective variances in equations (A19) and (A20).

$t\left(N-2,1-\frac{\gamma}{2}\right) = 100(1-\gamma)$ percentile of the *t*-distribution with $N-2$ degrees of freedom.

The estimates of the intercept and slope are correlated. Using the variances and the covariance of the estimates allows for obtaining the joint elliptical confidence region for these parameters. Description of the method of obtaining the joint confidence region is beyond the scope of this guideline.

The predicted bias (B_c) at a given medical decision level X_c is:

$$\hat{B}_c = a + (b-1)X_c \quad \text{(A24)}$$

Appendix A. (Continued)

The standard error for the bias can be calculated from the variances of intercept and slope, and their covariance (equations A19, A20, A21) as follows:

$$\hat{\sigma}_{\text{Bias}} = \sqrt{\hat{\sigma}_a^2 + X_c^2\hat{\sigma}_b^2 + 2X_c\hat{\sigma}_{ab}} \quad \text{(A25)}$$

The use of the above formulas for calculating $\hat{\sigma}_a$ and $\hat{\sigma}_b$ are not appropriate when the large sample approximation and other conditions mentioned in Miller are not satisfied.[1] The jackknife approach provided in Appendix K1 can be implemented under less restrictive conditions and is recommended in general situations. The bootstrap, repeatedly collecting *N* samples with replacement from the original samples, also provides a similarly less restrictive methodology to compute the standard errors.[3]

References for Appendix A

1 Miller RG Jr. *Beyond ANOVA, Basics of Applied Statistics.* New York, NY: Wiley; 1986:220-230.

2 Kendall M, Stuart A. *The Advanced Theory of Statistics, Volume 2: Inference and Relationship.* 4th ed. London, England: Griffin; 1979:406-407.

3 Davison AC, Hinkley DV. *Bootstrap Methods and their Application.* Cambridge, UK: Cambridge University Press; 1997.

Appendix B. Weighted Deming Regression

Abbreviations for Appendix B

CV coefficient of variation
SD standard deviation

For measurement procedures with extremely wide measuring intervals, the analytical SD is seldom constant. Rather, a proportional relationship might apply. In a situation in which proportional analytical errors for the measurement procedures are compared, the optimal approach is a weighted form of Deming regression analysis that takes into account the relationship between random error and measurand concentration. For a given sample measured by two analytical measurement procedures, *X* and *Y*:

$$x_i = X_{\text{Target}_i} + \varepsilon_{X_i} \tag{B1}$$

$$y_i = Y_{\text{Target}_i} + \varepsilon_{Y_i} \tag{B2}$$

$$Y_{\text{Target}_i} = \alpha + \beta X_{\text{Target}_i} \tag{B3}$$

x_i and y_i are the measured values, X_{Target_i} and Y_{Target_i} are the corresponding target values, ε_{X_i} and ε_{Y_i} are the random analytical error terms of the measurement procedures *X* and *Y*, α is the regression intercept, and β is the regression slope. The analytical SDs are assumed to be proportional to the target values (CV):

$$\sigma_X = \text{CV}_X X_{\text{Target}} \text{ and } \sigma_Y = \text{CV}_Y Y_{\text{Target}} \tag{B4}$$

Given a proportional relationship for the random errors, a weighted procedure assigns larger weights to measurements in the low range; the low-range measurements are more precise than measurements at higher concentrations that are subject to larger random errors. More specifically, distances from (x_i, y_i) to the line are inversely weighted according to the squared analytical SDs (variances) at a given concentration that express the random error. The regression line is then estimated so that the sum of squared weighted differences is minimized. The regression procedure is most conveniently performed using dedicated software. The principle of the computations is outlined below. Weighted averages, weighted sums of squares, and a weighted cross product are computed:

(B5, B6)

$$\bar{X}_w = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i} \qquad \bar{Y}_w = \frac{\sum_{i=1}^{N} w_i y_i}{\sum_{i=1}^{N} w_i}$$

$$u_w = \sum_{i=1}^{N} w_i (x_i - \bar{X}_w)^2 \tag{B7}$$

Appendix B. (Continued)

$$q_w = \sum_{i=1}^{N} w_i (y_i - \bar{Y}_w)^2 \tag{B8}$$

$$p_w = \sum_{i=1}^{N} w_i (x_i - \bar{X}_w)(y_i - \bar{Y}_w) \tag{B9}$$

The slope and intercept are estimated as[1,2]:

$$b = \frac{(\lambda q_w - u_w) + \sqrt{(u_w - \lambda q_w)^2 + 4\lambda p^{2}{}_w}}{2\lambda p_w} \tag{B10}$$

$$a_0 = \bar{Y}_w - b\bar{X}_w \tag{B11}$$

The weights are obtained by an iterative approach:

$$z_i = \frac{\hat{\mathrm{X}}_{\mathrm{Target}_i} + \lambda \hat{\mathrm{Y}}_{\mathrm{Target}_i}}{1 + \lambda}, \text{and} \tag{B12}$$

$$w_i = 1/[h(z_i)]^2, \tag{B13}$$

in which z_i estimates the true concentration of target *i*, and *h* is the fitted variance model (see Appendix J) of the variance. In Linnet's original proposal of constant CV, $h(z_i) = z_i$,[1]

$$W_i = \frac{1}{\left(\frac{\hat{\mathrm{X}}_{\mathrm{Target}_i} + \lambda \hat{\mathrm{Y}}_{\mathrm{Target}_i}}{1 + \lambda}\right)^2}, \tag{B14}$$

but other variance models can be used when appropriate.[1,2] It is here presumed that the ratio λ between the squared SDs (variances) for the random error components is constant throughout the measuring interval.

$$\lambda = {\sigma_x^2}\Big/{\sigma_y^2} = {\mathrm{var}(\varepsilon_X)}\Big/{\mathrm{var}(\varepsilon_Y)} \tag{B15}$$

λ can be based on the analytical CVs obtained from quality control results, for example. Otherwise, λ can, as default, be assigned the value 1. Without any knowledge of the ratio, for some purposes, the ratio can be varied to assess the sensitivity of the Deming regression to its value.

Appendix B. (Continued)

Bias at medical decision level(s) is calculated based on the estimates of slope and intercept. The jackknife approach provided in Appendix K1 can be used to calculate standard errors of regression parameters and SE of bias.

References for Appendix B

[1] Linnet K. Estimation of the linear relationship between the measurements of two methods with proportional errors. *Stat Med.* 1990;9(12):1463-1473.

[2] Linnet K. Evaluation of regression procedures for methods comparison studies. *Clin Chem.* 1993;39(3):424-432.