

Andrea Konnert,<sup>1,2,3</sup> Christoph Berding,<sup>3</sup> Sabine Arends,<sup>3</sup> Curt Parvin,<sup>4</sup> Curt L. Rohlfing,<sup>5</sup> Hsiao-Mei Wiedmeyer,<sup>4</sup> Randie Little,<sup>4</sup> Carla Siebelder,<sup>6</sup> and Cas Weykamp<sup>6</sup>

## Statistical Rules for Laboratory Networks

**ABSTRACT:** Within the definition of reference methods, laboratory networks and interlaboratory studies play an important role. To maintain the quality of the results statistical rules for quality control need to be defined. In this note we review statistical rules for the evaluation of laboratories participating in interlaboratory studies as well as data evaluation rules for the calculation of consensus means. The practicality of the derived rules is elaborated for a number of recent HbA1c interlaboratory studies

**KEYWORDS:** HbA1c reference standardization, interlaboratory studies, statistical rules, quality control, statistics for networks

### Introduction and Background

In the European Economical Area, assays in biomedical laboratories are regulated by a European law, the IVD Directive 98/79/EC. In this directive essential requirements for in vitro medical devices are described in general terms; details how to comply with the requirements are laid down in several standards [1,2]. A major specified requirement, directly related to standardization, is metrological traceability: when available, the calibration of biomedical assays must be traceable to reference measurement procedures and/or reference materials of a higher metrological order.

The measurement of hemoglobin A1c (HbA1c) in human blood is the most important biomedical marker for long-term assessment of the glycemic status in patients with diabetes mellitus, and goals for therapy are set at specific HbA1c target values [3,4]. The International Federation of Clinical Chemistry (IFCC) recognized the need for a reliable anchor for this major biomedical analyte and installed the IFCC Working Group on HbA1c standardization [5]. This group succeeded to develop a reference system of higher metrological order which has been approved by all member national societies of the IFCC [6].

The components of the reference system cover the upper part of the traceability chain; HbA1c is defined on the basis of its molecular structure, the primary reference measurement procedure is weighing of the pure analytes, the primary calibrator is a set of mixtures of pure HbA1c and pure HbA0, the secondary reference measurement procedure is the approved reference method and the secondary calibrators are whole blood panels to which values have been assigned with the reference method [6].

To ensure robustness of the reference system, the novel concept of a network of reference laboratories, rather than stand-alone ref-

erence laboratories, has been implemented. This concept guarantees continuity (if some labs discontinue, service is still available) and quality (e.g., ongoing re-approval of network laboratories, batch management of calibrators). Typical tasks of the interlaboratory studies are: (a) stability checks and approval of new calibrator batches, (b) re-approval of network laboratories periodically and approval of candidate network laboratories, and (c) value assignment to the provided third parties materials, e.g., calibrator and controls.

These tasks imply the calculation and application of data evaluation criteria based on sound statistical rules. This paper describes the statistical framework which eventually has been implemented in a set of ready-to-use SAS macros. Traceability and uncertainty calculations have been based on ISO 17511/18153 and GUM [7]. The HbA1c reference system is the first to use the network concept but will be applied for many more analytes in future. As such the model developed for HbA1c will also have a more general application.

### Design of the IFCC Interlaboratory Studies

Each year two interlaboratory studies are launched within the HbA1c network. The network laboratories base their calibration on approved calibrators supplied by the network. Within each study different types of samples are encountered: (a) samples for internal control (with known HbA1c concentration), (b) intercomparison samples (unknown HbA1c; basis for approval of labs), (c) third party samples (to which the network has to assign values), (d) calibrators of new lots (encountered to be checked and approved), and (e) calibrators of old approved lots (encountered to check long-term stability of the network).

In addition to the laboratories being already members of the network, there are candidate laboratories that want to join the network. To evaluate their performance, these laboratories measure control and intercomparison samples, too.

Within each lab each sample is processed by two enzymatic digests and two replications are executed per digest. The general design is illustrated by Table 1.

Table 1 exhibits the general design involving many samples of different types which are operated in a multitude of labs in two en-

Manuscript received February 21, 2005; accepted for publication August 8, 2005; published December 2005.

<sup>1</sup>Corresponding author, Tel. +49-8856-60 4647, e-mail: andrea.konnert@roche.com.

<sup>2</sup>Department of Statistics, University of Dortmund, Dortmund, Germany.

<sup>3</sup>Department of Biostatistics, Roche Diagnostics GmbH, Penzberg, Germany.

<sup>4</sup>Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO.

<sup>5</sup>University of Missouri School of Medicine, Columbia, MO.

<sup>6</sup>Queen Beatrix Hospital, Winterswijk, The Netherlands.

TABLE 1—Experimental design of interlaboratory study.

Study		Study A			
Sample Type	Lab	Lab 1		Lab 2	
	Sample	Digest 1	Digest 2	Digest 1	Digest 2
Controls	C1	Rep 1	Rep 1	Rep 1	Rep 1
		Rep 2	Rep 2	Rep 2	Rep 2
Intercomparison Sample	ICS1	Rep 1	Rep 1	Rep 1	Rep 1
		Rep 2	Rep 2	Rep 2	Rep 2

zomatic digests with two replications. For the sake of simplicity only a fraction of samples and labs is shown.

### Statistical Evaluation of the IFCC Interlaboratory Studies

The data of an IFCC interlaboratory study serve the following purposes: They allow for an approval of the data delivering laboratories. They provide reliable and accurate estimates for the analyte content of unknown samples, e.g., the intercomparison samples. Based on these estimates the newly produced network calibrators are verified. Accordingly the following statistical tasks have to be addressed: (i) The definition of statistical rules to approve laboratories (and candidate laboratories) as reliable data deliverers. (ii) Data of each sample must be checked for extreme results to obtain an accurate assigned value. In accordance with these steps two statistical rules have been derived which are applied consecutively. Figure 1 gives an illustration of the process flow.

As can be seen from Fig. 1, first the lab approval rule is applied to all laboratories. This rule is based on all intercomparison and calibrator samples. In a second step the data from the remaining approved labs are investigated for each sample specifically.

For each sample there are data from different labs, where each lab has performed two digests and two repetitions on each sample.

There could be a whole lab that reported extreme measurements in comparison to the other labs. An individual digest might go wrong, so the data reported from this digest might be very different from the data of the other digest. Finally we might observe an individual measurement within one digest which is extreme in comparison to the other results. The assignment data approval rule focuses on extreme labs and extreme observations within one lab neglecting the digest factor which is confounded with repetition by design.

Based on the concept that all data passed the data evaluation step, the assigned value of the sample and its uncertainty is derived. The assigned value is calculated as a straightforward mean, the uncertainty by taking into account the hierarchical structure of the study design.

### Evaluation of Laboratories

Mandel in his 1995 paper [8] has addressed the question of evaluating labs in interlaboratory studies by means of all sample results available. In particular, he suggested the investigation of lab-specific constant and proportional bias against the study average. Technically this can be done by a regression of the lab's sample means against the study sample means. In fact, this allows for an evaluation of the lab's individual position with regard to lab average. However, in order to derive general acceptance criteria valid

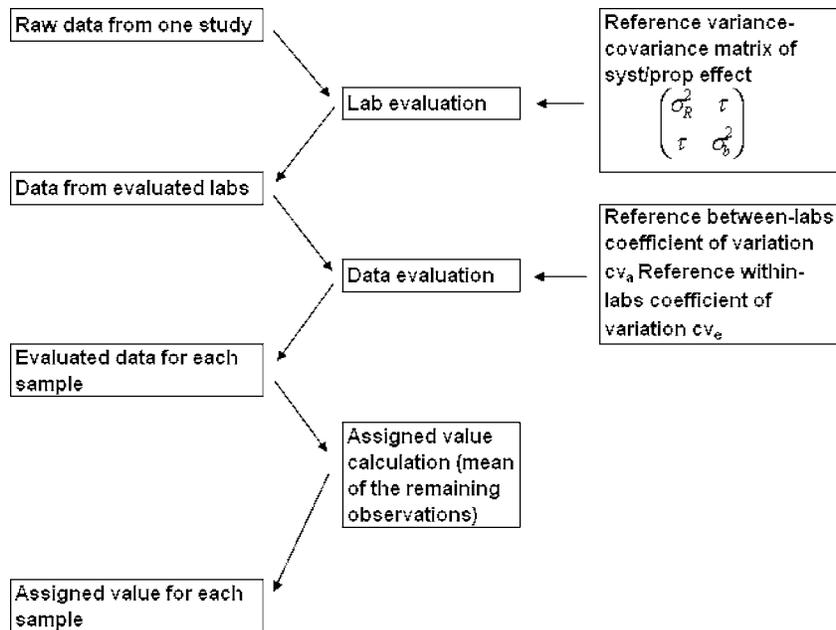


FIG. 1—The lab and data evaluation process flow. The figure provides a schematic picture of the process flow. The evaluation of the interlaboratory study data takes place in a two-step process focusing on (i) the evaluation of the labs, and (ii) the evaluation of the set of value assignment data. The scheme indicates the entry of quality characteristics from pilot process steps, too.

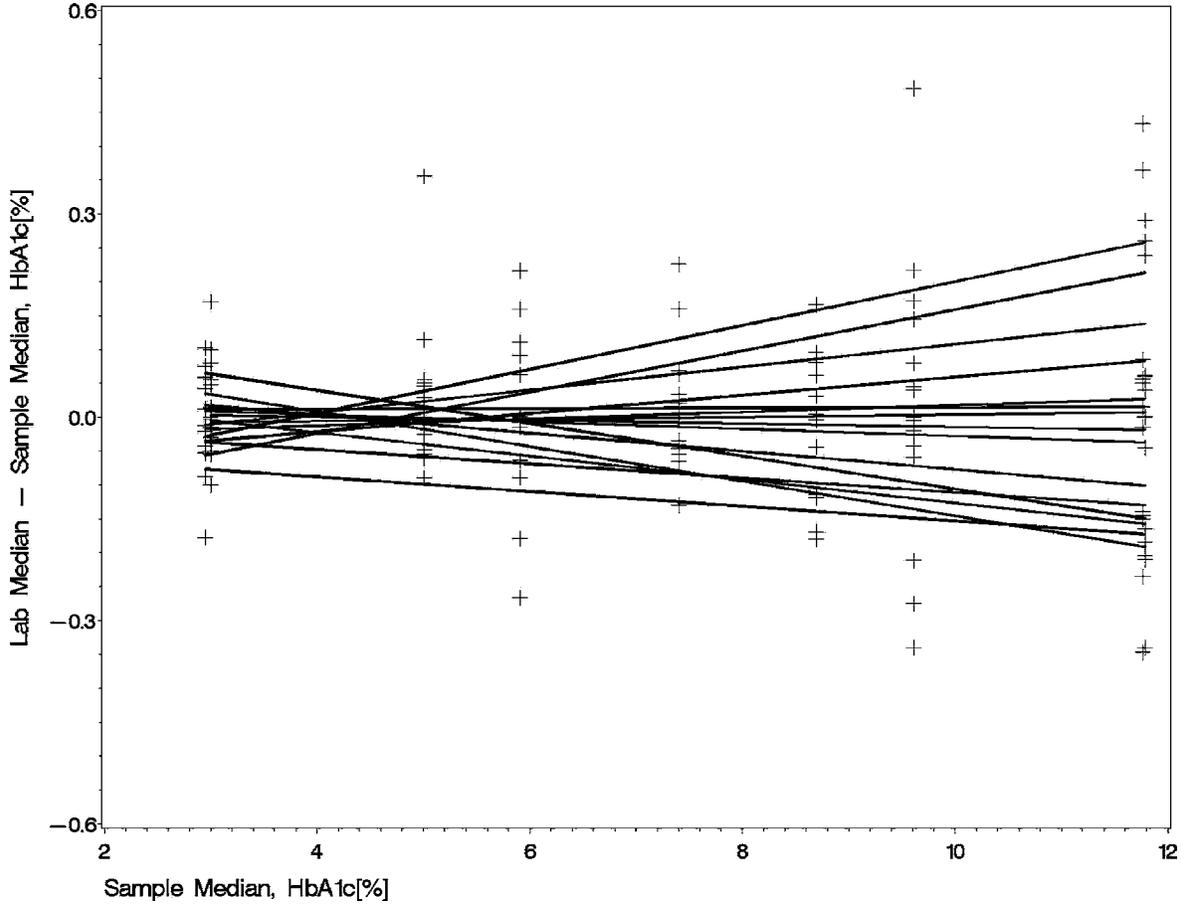


FIG. 2—Regression lines for each lab after fit of the random coefficient model. The figure illustrates the regression fits of the differences between the study lab medians minus the study sample medians plotted versus the study sample medians. The average lab regression line coincides with the abscissa. Labs with regression lines showing an angle or an offset to the abscissa are considered as being different with regard to a proportional and/or systematic bias, respectively.

for any future study, we make use of variance estimates for lab constant and/or proportional bias from a selected set of typical studies. If in comparison to average the observed constant and/or proportional bias of the lab under consideration is extreme, this lab fails the lab-evaluation rules.

The appropriate statistical model that reflects these ideas is a regression model with random effects. To introduce robustness against extreme observations either within one lab or a whole group of data the sample medians within one lab against the overall sample median is regarded, instead of the means. Furthermore, to achieve that the random effects have mean zero the differences between lab median  $\bar{Y}_{ij}$  and overall sample medians  $C_j$  are regressed on the overall sample medians  $C_j$ . The model can be written as

$$\bar{Y}_{ij} - C_j = R_i + b_i C_j + \varepsilon_{ij} \quad i = 1, \dots, I, j = 1, \dots, J$$

Here  $R_i$  — the constant bias — is modeled as a normal distributed random variable with mean zero and variance  $\sigma_R^2$ , and  $b_i$  — the proportional bias — is a normal distributed random variable with mean zero and variance  $\sigma_b^2$ . We also have to take into account the correlation between constant and proportional bias: the variance-covariance matrix of these two random variables is given by

$$\text{Var} \begin{pmatrix} R_i \\ b_i \end{pmatrix} = \Sigma = \begin{pmatrix} \sigma_R^2 & \tau \\ \tau & \sigma_b^2 \end{pmatrix}$$

Furthermore, the residuals are assumed normally distributed independent random variables with mean zero and variance  $\sigma_\varepsilon^2$ .

This model can be fitted by maximum-likelihood methods [9], leading to estimates of the variance-covariance matrix  $\Sigma$  of constant and proportional bias and the residual variance  $\sigma_\varepsilon^2$ .

Furthermore, for each lab the pair of constant and proportional bias is predicted. Figure 2 illustrates the fitted regression lines for each lab for a specific HbA1c interlaboratory study

Based on the variance-covariance matrix of the lab effects, we could define an acceptance region by the joint confidence interval of constant and proportional bias to the joint  $1 - \alpha$  confidence level,  $0 < \alpha < 1$ . It is of the form

$$I = \left\{ \begin{pmatrix} R \\ b \end{pmatrix} \left| \frac{1}{\sigma_R^2 \sigma_b^2 - \tau^2} \cdot (R^2 \sigma_b^2 - 2\tau \cdot R \cdot b + b^2 \sigma_R^2) \leq q_{1-\alpha, J} \right. \right\}$$

with  $q_{1-\alpha, J}$  denoting the  $1 - \alpha, J$  quantile of the chi-square distribution with two degrees of freedom and  $\alpha, J = 1 - (1 - \alpha)^{1/J}$ . For the lab evaluation rule of the HbA1c network  $\alpha$  was taken as  $\alpha = 0.01$ . In order to bring this expression to work we need the entries of the variance-covariance matrix. As mentioned previously, in order to have general rules we take these estimates from a set of specifically selected pilot studies.

We derive them empirically from the six selected pilot studies. For each study the measured results of the calibrators (except zero level calibrators) and intercomparison samples are used for the cal-

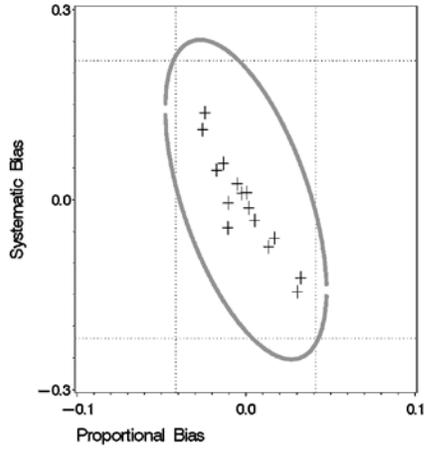


FIG. 3—Evaluation region for lab approval and regression parameters of the constant and proportional bias of an interlaboratory study. The figure illustrates the regression coefficients for a study under evaluation together with the acceptance region as derived from the selected pilot studies. Within the fit of the random coefficient model for each lab of the study under evaluation a constant and proportional bias has been determined. As is well known from regression theory there is a negative correlation between the proportional and the constant bias. As all points lie within the acceptance region all labs of this study have been approved.

culuation. For each of the selected pilot studies a separate random coefficient model is fitted by maximum likelihood estimation leading to six estimators of the variance-covariance matrix of the constant and proportional bias. The reference variance-covariance matrix is defined by the element-wise median of the six matrices.

A graphical representation of the lab evaluation region together with a respective set of lab data under evaluation is shown in Fig. 3.

After the laboratories have been approved as valuable data providers, the datasets for individual samples used for value assignment have to be checked. This is addressed in the next section.

## Evaluation of Value Assignment Data

Precedent to calculating any assigned value, the value assignment data need to be approved. This means that data from each sample are examined individually for the presence of (i) extreme labs, i.e., labs where all observations are extreme in relation to the other labs for the individual sample, and (ii) individual extreme data points within one lab (see also [10,11]). The detected extreme observations have to be excluded and then the consensus mean of the remaining data points becomes the assigned value of the sample under consideration.

To set up the statistical model recall once again the measurement design of the HbA1c samples: samples are measured in different labs, each lab accomplishes two digestions and measures two repetitions per digestion. Based on this structure the results from one sample can be modeled by a One-Way Classification Model with random effects. Note that based on empirical evidence the digestion factor is not modeled explicitly.

The One-Way Classification Model with random effects is of the form for all  $j=1, \dots, J$

$$Y_{ik}^j = \mu + a_i + \varepsilon_{ik}, \quad i = 1, \dots, I, \quad k = 1, \dots, K_i$$

$Y_{ik}^j$  denotes the measured result from sample  $j$ , measured in lab  $i$  and repetition  $k$ . The index  $j$  is written as a superscript to indicate that data from each sample are analyzed independently. Here  $a_i$  is

the lab effect of lab  $i$ , modeled as a normally distributed random variable with mean 0 and variance  $\sigma_a^2$ —the between-labs variance. The error  $\varepsilon_{ik}$  are normally distributed random variables with mean 0 and variance  $\sigma_\varepsilon^2$ —the within-labs variance. Furthermore, it is assumed that all random variables are independent.

The key idea is to derive limits for the residuals and the lab effects. Any residual or lab effect outside pre-specified marginal quantiles is considered as extreme. To be able to apply the (same) limits to a multitude of standardization actions, the estimation of the between-labs and within-labs variance should not depend on the actual data, but should be estimated in advance. As this reference we use once again the intercomparison and calibrators samples of a selected set of pilot studies. Data from candidate labs are not included in the calculation as well as calibrators with zero level.

For all remaining samples we calculate the between-labs variance  $\sigma_{a_j}^2$  and between-labs standard deviation  $\sigma_{a_j}$  as well as the within-labs variance and within-labs  $\sigma_{\varepsilon_j}^2$  standard deviation  $\sigma_{\varepsilon_j}$ . Moreover we note that the standard deviations depend linear on the concentration and calculate a between-labs coefficient of variation  $cv_a$  and a within-labs coefficient of variation  $cv_\varepsilon$  by means of the linear models

$$\sigma_{\varepsilon_j} = cv_\varepsilon \cdot \mu_j + \varepsilon_j \quad \text{and} \quad \sigma_{a_j} = cv_a \cdot \mu_j + \varepsilon_j.$$

Thus, to account for concentration dependence of the between-lab and within-lab variances the limits for the residuals and the lab effects are derived from a reference between-labs coefficient of variation and a reference within-labs coefficient of variation rather than from the respective variances.

The location parameter  $\mu$  and the lab random effects  $a_i$  are derived from the actual data under consideration. In particular for each sample  $\mu$  is derived in a robust way as the median of the lab medians

$$\hat{\mu} = \underset{i}{\text{med}}(\underset{k}{\text{med}}(y_{1k}), \underset{k}{\text{med}}(y_{2k}), \dots, \underset{k}{\text{med}}(y_{Ik}))$$

The  $i$ th lab effect is estimated by

$$\hat{a}_i = \frac{J\tilde{\sigma}_a^2}{\tilde{\sigma}_\varepsilon^2 + J\tilde{\sigma}_a^2}(\text{med}_i - \hat{\mu})$$

according to the formula of the best linear predictor. Note that the estimated lab effects depend on the variance components  $\tilde{\sigma}_a(cv_a, \mu)$  and  $\tilde{\sigma}_\varepsilon(cv_\varepsilon, \mu)$  as calculated from the selected pilot studies.

Finally the within-lab residuals are given by

$$\hat{\varepsilon}_{ij} = |y_{ik} - \hat{\mu} - \hat{a}_i|$$

Figure 4 shows the derivation of the reference coefficients of variation for the between lab and the within-lab situation based on a set of selected pilot studies

Table 2 gives the reference between-lab and the within-lab coefficients of variation for the study data shown in Fig. 4.

We define a lab as extreme within the set of labs under consideration, if the random effect  $a_i$  lies outside the set

$$\text{in}_L(\alpha, \sigma_a) = \{a: |a| \leq z_{1-\alpha/2} \cdot \tilde{\sigma}_a(cv_a, \mu)\}$$

where  $z_{1-\alpha/2}$  denotes  $1-\alpha/2$  quantile of the standard normal distribution with  $\alpha=0.01$  and  $\alpha_j=1-(1-\alpha)^{1/I}$  with  $I$  denoting the number of labs.

Figure 5 shows the lab effects for an actual HbA1c interlaboratory study for each sample. The inlier region is indicated by hori-

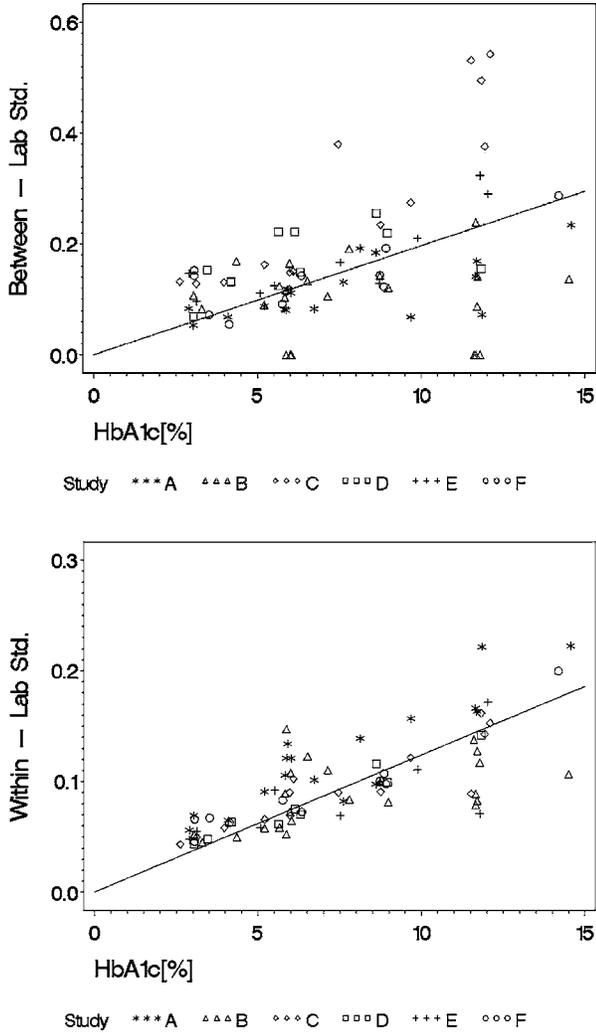


FIG. 4—The between-lab and within-lab standard deviation versus sample mean and fitted regression line. A least square linear regression of between-labs standard deviation and the within-labs standard deviation versus sample mean has been executed. The respective slopes provide an estimate for the between-labs and the within-labs coefficient of variation, respectively. They are used in the approval of the value assignment data to check for extreme labs and individual measurements. The symbols indicate the different studies in which the particular sample was measured.

zonal bars. According to the constant coefficient of variation model the absolute width of this region increases with concentration.

The variance components  $\tilde{\sigma}_a(cv_a, \mu)$  and  $\tilde{\sigma}_e(cv_e, \mu)$  are taken from the selected pilot studies (see Fig. 4 and Table 2).

We define an observation  $y_{ij}$  as an  $\alpha$ -extreme value within the  $i$ th lab, if it lies outside the region

$$in_L(\alpha, \sigma_e) = \{y: |y - \mu - a_i| \leq z_{1-\alpha_N/2} \cdot \tilde{\sigma}_e(cv_e, \mu)\}$$

where  $z_{1-\alpha_N/2}$  denotes the  $1-\alpha_N/2$  quantile of the standard normal distribution with  $\alpha=0.01$  and  $\alpha_N=1-(1-\alpha)^{1/N}$  where  $N=\sum_i K_i$  denotes the number of measurements. Again the variance components

TABLE 2—Reference coefficients of variation.

Reference between-labs CV $cv_a$	Reference within-labs CV $cv_e$
0.02	0.0125

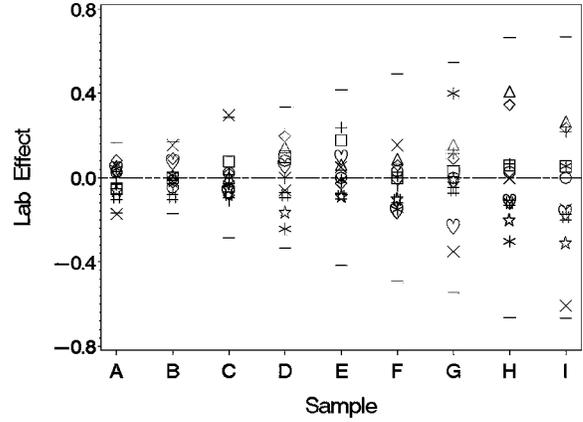


FIG. 5—Between-lab effects of the study under consideration and the respective acceptance limits. The figure shows the lab effects for each sample of the study sorted by sample concentration. Horizontal bars at the upper and the lower end of the vertically displayed lab results indicate the respective acceptance limits. All lab effects within these bars are accepted for value assignment calculations.

$\tilde{\sigma}_a(cv_a, \mu)$  and  $\tilde{\sigma}_e(cv_e, \mu)$  are taken from the selected pilot studies (see Fig. 4 and Table 2).

Figure 6 shows the individual within-lab residuals for an actual HbA1c interlaboratory study for each sample. The inlier region is indicated by horizontal bars. According to the constant coefficient of variation model the absolute width of this region increases with concentration.

After the lab and data approval steps have been completed, assigned values of the samples can be calculated based on the remaining data.

### Assigned Value and Uncertainty Calculation

Finally, after the lab and data evaluation steps have been completed, we are left with data which comply to the quality requirements of the HbA1c standardization network. Accordingly these results enter the subsequent value assignment calculation and the arith-

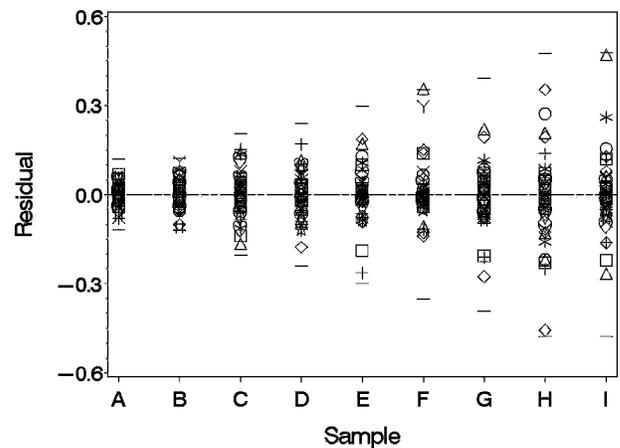


FIG. 6—Within-lab effects of the study and the respective acceptance limits. The figure shows the within-lab residuals of each sample of the study sorted by sample concentration. Horizontal bars at the upper and the lower end of the vertically displayed sample residuals indicate the respective acceptance limits. All data within these bars are accepted for value assignment calculations.

metric mean of the measured results of a given sample becomes the new assigned value

$$c_{av} = \bar{Y}^j = \frac{1}{N} \sum_{i,k} Y_{ik}^j$$

Two components contribute to the uncertainty of the assigned value: (i) errors in the measurement process due to the between-labs and within-labs variances, and (ii) errors in the calibration due to calibrator uncertainties. Recall that the calibrators are samples with a weighed amount of HbA1c and HbA0 [6] that are adjusted for impurities. The variance of the assigned value of a sample thus takes the form

$$Var(c_{av}) = Var_{me}(c_{av}) + Var_{cal}(c_{av})$$

The first term on the right-hand side is calculated using standard analysis of variance techniques for nested designs. Recall that each sample is measured in different labs, in two digests per lab and two repetitions per digest. However, after the lab and data evaluation due to the removal of labs or individual measurements the design of a single sample may not be balanced anymore.

We can model the remaining data of a sample once again as a One-Way Classification model with random effects,

$$Y_{ik}^j = \mu + a_i + \varepsilon_{ik}, \quad i = 1, \dots, I, \quad k = 1, \dots, K_i$$

The variance of a single measurement is given by the sum of between-labs variance and within-labs variance, i.e.,

$$Var_{me}(Y_{ik}^j) = \sigma_a^2 + \sigma_\varepsilon^2$$

As the measurements within one lab are correlated with  $Cov(Y_{ik}^j, Y_{ik'}^j) = \sigma_a^2$ , the variance of the mean of the measurements reads

$$Var_{me}(c_{av} = \bar{Y}^j) = \frac{1}{I} \sigma_a^2 + \frac{1}{N} \sigma_\varepsilon^2$$

where  $N = \sum_i K_i$ .

Finally, the between-labs and within-labs variances of the data under consideration need to be estimated. Using the ANOVA estimators [12] we obtain

$$\hat{\sigma}_a^2 = \frac{MSA - MSE}{\left(N - \sum_i K_i^2/N\right)/(I-1)}, \quad \hat{\sigma}_\varepsilon^2 = MSE$$

where

$$MSE = \frac{1}{N-I} \sum_{i,k} Y_{ik}^j{}^2 - \sum_i K_i \cdot \bar{Y}_i^j{}^2,$$

$$MSA = \frac{1}{I-1} \sum_i K_i \cdot \bar{Y}_i^j{}^2 - N \cdot \bar{Y}^j{}^2.$$

The variance component of the assigned value due to the calibration process addresses potential errors of calibrator production and value assignment. Presuming that these occur independently and for a two-point linear calibration with respective calibrators bracketing the value assignment sample we may use Gaussian error propagation [13] to arrive at the following expression

$$Var_{cal}(c_{av}) = \langle s \rangle^2 \cdot Var(c_{cal,i}) + [1 - \langle s \rangle]^2 \cdot Var(c_{cal,i+1})$$

where  $\langle s \rangle = (c_{cal,i+1} - c_{av}) / (c_{cal,i+1} - c_{cal,i})$  and  $c_{cal,i}, c_{cal,i+1}$  denote the assigned values of the calibrators with  $c_{cal,i} \leq c_{av} < c_{cal,i+1}$  and

$Var(c_{cal,i}), Var(c_{cal,i+1})$  the variances of these calibrators resulting from production. Strictly speaking, for individual value assignments calibrator value errors provide a bias rather than a random error term. However, in order to stay within the GUM framework [7] we model this term as a variance. The details of the derivation of the production variances will be considered elsewhere.

Taking all expressions together we finally obtain the uncertainty of the assigned value of sample  $j$  by

$$u(c_{av}) = \sqrt{Var_{me}(c_{av}) + Var_{cal}(c_{av})}$$

## Conclusion

In this paper we presented statistical rules for the evaluation of data from interlaboratory studies. In particular these rules have been developed to match the needs of the IFCC network on HbA1c standardization. Progressing from the large to the small, the rules start with an evaluation of the lab performance (constant and proportional bias) based on all data and a more detailed investigation of the individual value assignment data for lab as well as individual measurement results.

For each lab a constant and proportional bias is calculated based on the measured results of all samples. If the bias pair lies outside a predefined confidence region the lab is not approved as a reliable data deliverer. After the lab evaluation, data for each sample are analyzed individually. Within these data we seek for extreme labs in comparison to the other labs as well as for extreme single measurements within one lab. The necessary input for appropriate reference criteria has been provided from six previous (pilot) studies. Finally, a straight consensus mean of the approved data gives the new assigned value and its uncertainty is calculated.

As an example, the procedures outlined have been applied to the data from HbA1c IFCC standardization studies. The results are well in line with evaluation by the laboratory experts. Indeed, the system of statistical rules allows for an automatic and objective processing of the study data thus strongly supporting the practical operation of the network. Backed by these observations we are confident that the statistical rules of the type presented here might also be very valuable in other interlaboratory studies, too.

## Acknowledgment

This work has been done in collaboration with the IFCC Working Group on Standardization of HbA1c, comprising of: Dr. Kor Miedema, Isala Klinieken, Zwolle, The Netherlands (Chair), Prof. Andrea Mosca, Università degli Studi di Milano, Milan, Italy (Secretary), Dr. Cas Weykamp, Queen Beatrix Hospital, Winterswijk, The Netherlands (Network Coordinator), Randie Little, Ph.D., University of Missouri School of Medicine, Columbia MO, David Sacks, M.D., Brigham & Women's Hospital, Boston MA, Dr. Ian Goodall, A&RMC, Austin Campus, Heidelberg, VA, Australia, Tadao Hoshino, M.D., Ph.D., Institute of Biopathological Medicine, Kanagawa, Japan, Jan-Olof Jeppsson, M.D., Ph.D., Malmö University Hospital, Malmö, Sweden, Dr. Gary John, Norfolk and Norwich University Hospital, Norwich, UK, Gary Myers, Ph.D., Centers for Disease Control and Prevention, Atlanta, GA.

## References

- [1] EN ISO 17511: In-vitro diagnostic medical devices—Measurement of quantities in biological samples—Metrological traceability of values assigned to calibrators and control materials (ISO/FDIS 17511:2002).
- [2] EN ISO 18153: In-vitro diagnostic medical devices—Measurement of quantities in samples of biological origin—Metrological traceability of values for catalytic concentration of enzymes assigned to calibrators and control materials (ISO/DIS 18153:2000).
- [3] The Diabetes Control and Complications Trial Research Group, “The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-term Complications in Insulin-dependent Diabetes Mellitus,” *N. Engl. J. Med.*, Vol. 329, 1993, pp. 977–986.
- [4] ADA., Clinical practice guidelines 2001, *Diabetes Care*, 2001, Vol. 24, (Suppl 1): S1–133.
- [5] Hoelzel, W. and Miedema, K., “Development of a Reference System for the International Standardization of HbA1c/ Glychemoglobin Determinations,” *J. Int. Fed. Clin. Chem.*, Vol. 9, 1996, pp. 62–67.
- [6] Jeppsson, J.-O., Kobold, U., Barr, J., Finke, A., Hoelzel, W., Hoshino, T., et al., “Approved IFCC Reference Method for the Measurement of HbA1c in Human Blood,” *Clin. Chem. Lab. Med.*, Vol. 40, 2002, pp. 78–89.
- [7] *Guide to the Expression of Uncertainty in Measurement*, ISBN 92-67-10188-9, ISO, Geneva 1995.
- [8] Mandel, J., “Structure and Outliers in Interlaboratory Studies,” *J. Test. Eval.* Vol. 23, No. 5, 1995, pp. 364–369.
- [9] Longford, N. T., *Random Coefficient Models*, Clarendon Press, Oxford, 1993.
- [10] Gather, U. and Wellmann, J., “Identification of Outliers in a One-Way Random Effects Model,” Technical Report, University of Dortmund, 2000.
- [11] Konnert, A., “Linear Random Effects Models with Applications in Clinical Chemistry,” Diploma Thesis, University of Ulm, 2003.
- [12] Searle, S. R., Casella, G., and McCulloch, Ch. E., *Variance Components*, John Wiley & Sons, Inc., New York, 1992.
- [13] Dietrich, C. F., *Uncertainty, Calibration and Probability*, Adam Hilger, Bristol, 1991.