Kiyoshi Ichihara*, Yesim Ozarda, George Klee, Joely Straseski, Nikola Baumann and Kiyohide Ishikura on behalf of the Committee on Reference Intervals and Decision Limits, International Federation for Clinical Chemistry and Laboratory Medicine

Utility of a panel of sera for the alignment of test results in the worldwide multicenter study on reference values

Abstract

Background: In a planned International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) worldwide study on reference intervals (RIs), a common panel of serum samples is to be measured by laboratories from different countries, and test results are to be compared through conversion using linear regression analysis. This report presents a validation study that was conducted in collaboration with four laboratories.

Methods: A panel composed of 80 sera was prepared from healthy individuals, and 45 commonly tested analytes (general chemistry, tumor markers, and hormones) were measured on two occasions 1 week apart in each laboratory. Reduced major-axis linear regression was used to convert reference limits (*LL* and *UL*). Precision was expressed as a ratio of the standard error of converted *LL* or *UL* to the standard deviation (SD) comprising RI (approx. 1/4 of the RI width corresponding to between-individual SD). The allowable and optimal levels of error for the SD ratio (SDR) were set as ≤0.250 and ≤0.125, respectively, in analogy to the common method of setting limits for analytical bias based on between-individual SD.

Results: The values for the calculated SDRs depended upon the distribution patterns of test results: skewness toward higher values makes SDR_{LL} lower and SDR_{UL} higher. However, the CV of the regression line slope, *CV*(*b*), is less affected by skewness. The average of SDR_{LL} and SDR_{UL} (aveSDR) correlates closely with *CV*(*b*) (r=0.995). The aveSDRs of ≤0.25 and ≤0.125 corresponds approximately to *CV*(*b*) values of ≤11% and ≤5.5%, respectively. For all results (i.e., n=80), conversion was allowable (optimal) in 98% (89%) of the analytes, as judged by *CV*(*b*). Resampling studies using random subsets of data with a data size (n) of 70 to 20 revealed that SDRs and *CV*(*b*) gradually increase with reduction of n, especially with n ≤30.

Conclusions: *CV*(*b*) is a robust estimator for judging the convertibility of reference values among laboratories, even with a skewed distribution. Assuming 40 sera to be a practical size for the panel, reference values of 89% (80%) of analytes examined were made comparable by regression analysis with the allowable (optimal) level of precision.

Keywords: method comparison; multicenter study; panel of sera; reduced major-axis regression; reference interval.

***Corresponding author: Kiyoshi Ichihara,** MD, PhD, Faculty of Health Sciences, Department of Clinical Laboratory Sciences, Graduate School of Medicine, Yamaguchi University, Minami-Kogushi 1-1-1, Ube 755-8505, Japan, Phone: +81-836-22-2884, Fax: +81-836-35-5213, E-mail: ichihara@yamaguchi-u.ac.jp
Yesim Ozarda: Department of Biochemistry and Clinical Biochemistry, Uludag University Medical School, Bursa, Turkey
George Klee and Nikola Baumann: Laboratory Medicine and Pathology, College of Medicine, Mayo Clinic, Rochester MN, USA
Joely Straseski: ARUP Laboratories, University of Utah School of Medicine, Salt Lake City, UT, USA
Kiyohide Ishikura: PhD, Beckman Coulter Japan, Tokyo, Japan

Abbreviations: AFP, α-fetoprotein; Alb, albumin; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AMY, amylase; AST, aspartate aminotransferase; C3, complement component 3; C4, complement component 4; Ca, total calcium; CA125, carbohydrate antigen 125; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; CK, creatine kinase; Cl, chloride; CRE, creatinine; CV(*b*), coefficient of variation of slope *b;* CV_A, analytical CV; CV_I, within-individual CV; DR, Deming regression; E2, estradiol; Fe, iron; Ferritin, ferritin; folate, folic acid; FSH, follicle-stimulating hormone; FT3, free triiodothyronine; FT4, free thyroxine; GGT, γ-glutamyltransferase; Glu, glucose; GM, grand mean; HDL-C, HDL cholesterol; IP, inorganic phosphate; K, potassium; LDH, lactate dehydrogenase; LDL-C, LDL cholesterol; LH, luteinizing hormone; *LL*, lower limit; MAR, reduced major axis regression; Na, sodium; PCR, principal component regression; PRL, prolactin;

PSA, prostate-specific antigen; PTH, intact parathyroid hormone; RI, reference interval; RMP, reference measurement procedure; SD_G, between-individual SD; SD_I, within-individual SD; SD_{RI}, SD comprising RI $\left(\sqrt{SD_I^2+SD_G^2}\right)$; SDR_{LL}, SD ratio of *LL; SDR_{UL}*, SD ratio of *UL*; SE, standard error; TBil, total bilirubin; TCho, total cholesterol; Tf, transferrin; TG, triglyceride; TP, protein, total; TSH, thyroid-stimulating hormone; TTR, transthyretin (prealbumin); UA, uric acid; *UL*, upper limit; UN, urea nitrogen; VB12, vitamin B12.

Introduction

Plans for a worldwide multicenter study on reference values were developed progressively since 2010 by the Committee on Reference Values and Decision Limits (C-RIDL) of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), with the participation of the USA, the UK, Turkey, Japan, China, Saudi Arabia, India, South Africa, and the Philippines. Its objectives are (1) to establish country-specific RIs from 500 or more healthy individuals in a harmonized manner using a common protocol (2) to make test results comparable across participating countries through the common measurement of a specified panel of sera on the basis of linear regression analysis, and (3) to explore sources of variation of the aligned test results using information obtained from a questionnaire given to each individual in the pool.

The detailed protocol, standard operating procedures (SOPs) and the questionnaire to be used for subject recruitment, sample collection, specimen processing, measurements, and data analysis have been discussed in the companion paper [1]. The schemes for implementing the adopted study were as follows:

1. Each country will conduct its own multicenter study to derive country-specific RIs and explore sources of variation of the test results relevant to that country. The target analytes and questionnaire items, which are not included in the general protocol, can be varied according to local needs for the survey.
2. Collaborating local laboratories in each country will recruit appropriate healthy volunteers and collect and process specimens as per the SOPs.
3. A centralized measurement scheme will be used to eliminate variation attributable to differences in analytical methods. One laboratory (or two) will act as a central laboratory in each country, receive specimens from local laboratories, and carry out collective measurement. The central laboratory can use any assay platform.
4. For standardized analytes, it is requested that the laboratory ensure that RIs are traceable to the RMPs for standardized analytes through the measurement of certified reference materials (CRMs) or value-assigned sera provided by reference laboratories listed by the Joint Committee for Traceability in Laboratory Medicine.
5. For non-standardized analytes, RIs determined centrally will be converted to those of each participating laboratory based on the results of local comparative measurements (cross-check testing) using aliquots of the specimens taken for the study.
6. For worldwide comparison, test results will be made comparable across the collaborating countries by the measurement of the same panel of sera by all laboratories.

Among these various schemes, the successful alignment of test results among countries is of utmost importance to allow the comparison of country- or region-specific reference intervals (RIs). This will depend largely on the reliability of the comparison of test results on the panel of sera.

In this study, a panel of sera from 80 healthy volunteers was freshly prepared and then distributed to four reference laboratories in Turkey, Japan, and the US (two sites). All samples were assayed twice in each laboratory, 1 week apart, for 45 commonly tested analytes (general chemistry and immunoassays). The practicability of converting test results or RIs from one laboratory to another was evaluated based on variability of the slope and converted reference limits as a function of the required sample size in the panel.

Materials and methods

Preparation of the panel of sera

Recruitment of healthy volunteers

With the approval of the Ethical Committee in Yamaguchi University Graduate School of Medicine, healthy volunteers were recruited mainly from the hospital's health workers. Only subjects who were between 20 and 65 years old and who were subjectively healthy were included in the study. Excluded were those who were anemic, overweight (BMI >30), known to have hyperlipidemia, known to be a carrier of HBV, HCV, or HIV, pregnant or had given birth within 1 year, heavy drinkers (ethanol >70 g/day on average), heavy smokers (≥25 cigarettes/day), or taking more than three drugs on a regular basis for a chronic disease.

A total of 80 healthy volunteers agreed to donate blood for producing the panel after the objectives, precautions about sampling of a large volume of blood, and benefits of participation (free testing of 45

analytes listed below) had been explained to them. Written consent was obtained from each volunteer before participation in the study.

Blood collection and specimen preparation

After overnight fasting, blood was drawn using 19-gauge butterfly needles and 55-mL disposable syringes with total draws of 220 mL from male subjects and 180 mL from female subjects. Each tube of blood was transferred into a large conical centrifuge tube with 45-mL capacity. Then, the conical tubes were left at room temperature for 60 min and centrifuged at room temperature for 10 min at 1200 *g*. The serum was poured into a glass beaker with 200-mL capacity. After thorough mixing, the serum, with a total volume of approximately 90 mL, was aliquoted with equal volumes of 1.5 mL into 1.8-mL capacity CryoTube cryovials (Nunc, Roskilde, Denmark). Thus, 60 or more serum aliquots were obtained from each volunteer. They were immediately stored in a freezer at −80°C. Single serum aliquots from each of the 80 individuals were assembled into 80-member panel sets, and 60 of these sets were placed into 60 individual freeze boxes for distribution to the participating sites.

Collaborating laboratories and target analytes

Four sets of samples were transported at −80°C to each of the following four collaborating laboratories: Central Clinical Laboratory of Mayo Clinic in Rochester, MN, USA; ARUP Laboratories in Salt Lake City, UT, USA; Clinical Laboratory of Uludag University Hospital, Bursa, Turkey; Beckman Coulter Japan's Mishima Central Laboratory, Mishima, Japan. Two sets were to be used to assay analytes by chemical assays and immunoturbidimetry and two for testing other analytes by immunoassay methods.

The following 45 analytes in each serum were measured: total protein (TP), albumin (Alb), urea nitrogen (UN), uric acid (UA), creatinine (CRE), total bilirubin (TBil), sodium (Na), potassium (K), chloride (Cl), calcium (Ca), inorganic phosphate (IP), iron (Fe), glucose (Glu), total cholesterol (TCho), triglycerides (TG), HDL-cholesterol (HDL-C), LDL cholesterol (LDL-C), AST, ALT, ALP, LDH, GGT, CK, amylase (AMY), complement components 3 and 4 (C3 and C4, respectively), transferrin (Tf), transthyretin (TTR), vitamin B12 (VB12), folate, ferritin, AFP, CEA, CA19-9, CA125, PSA, prolactin (PRL), luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol (E2), thyrotropin (TSH), free thyroxine (FT4), free triiodothyronine (FT3), intact parathyroid hormone (PTH), and insulin.

The assay platforms and methods used in each laboratory are as listed in Table 1.

Assay procedures and quality control

All 80 sera in the panel were separated into two to four parts and measured in singleton on two different days to evaluate the between-day variation. The measurements were repeated using the second set of samples 1 week later, using the same schedule, to determine the reproducibility of test results. During the assay, a minipanel, consisting of five deep-frozen sera prepared from five of the healthy volunteers, was measured each day in singleton to monitor the stability of the assay over the study period. In addition, to confirm the status of standardization of the assays in each laboratory, the following CRMs with assigned values, all supplied as sets of vials containing deeply frozen specimens, were purchased from the Reference Material Institute for Clinical Chemistry Standards (ReCCS), Japan: JCCRM224-5a for HDL-C, LDL-C, TCho, and TG consisted of one set of four reference materials to be measured in triplicate in a single day; JCCRM321-5 for UA, CRE, Glu, and urea consisted of two sets of three materials with different levels of these analytes to be measured on two separate days each in triplicate; JCCRM521-10 for Na, K, Cl, Ca, Mg, and IP consisted of three sets of two materials to be measured on three separate days, each day in triplicate.

Data analysis

Cross-comparison of test results among laboratories

There are a variety of ways for expressing the linear functional relationship of test results between any two laboratories [2–7]. Principal component regression (PCR) requires the assumption that the slope of the regression line equals 1.0 and the use of equivalent scales and units of measurement in both laboratories for use in method comparison, whereas reduced major axis regression (MAR) does not require these assumptions [2, 3]. Deming regression (DR) [7] overcomes the problem of PCR by introducing an error ratio (λ) of test results of one laboratory to the other. However, determining λ requires replicate measurements of each specimen in both laboratories. Therefore, the most appropriate and generally applicable method of regression for the purpose of this study was MAR [2]. It is also known as geometric mean regression [6] or least-square product regression [2] and is expressed as follows, with $\bar{y}$, $\bar{x}$, and b denoting the means of x (test results from laboratory X), y (corresponding results from laboratory Y), and the slope, respectively:

$$y=\bar{y}+b\left(x-\bar{x}\right) \tag{1}$$

Slope b is determined by the following formula:

$$b=\sqrt{\frac{S_{yy}}{S_{xx}}}\quad\left(S_{xx}=\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2,\ S_{yy}=\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2\right), \tag{2}$$

where n, S_{xx}, and S_{yy} denote the sample size and sums of squared deviations of values from laboratories X and Y, respectively.

Using the regression equation (1), the lower limit (LL_0) and upper limit (UL_0) of an RI derived using test results (x) of laboratory X can be converted to those of laboratory Y (LL and UL) using the following formulae:

$$LL=\bar{y}+b\times\left(LL_0-\bar{x}\right),\quad UL=\bar{y}+b\times\left(UL_0-\bar{x}\right).$$

The precision of the conversion can be evaluated from the magnitude of standard error (SE) of the slope b [$SE(b)$], LL [$SE(LL)$], or UL [$SE(UL)$], where the SE in this context implies standard deviation (SD) of any summary value X (statistics, such as b, r, LL, UL).

Assuming equal variance of the data points (homoscedasticity) around the regression line for the range of regression and an equal ratio of inherent errors on both x and y, $SE(b)$ can be computed by the following formula using Pearson correlation coefficient r and the data size n. This formula has not been described in the past

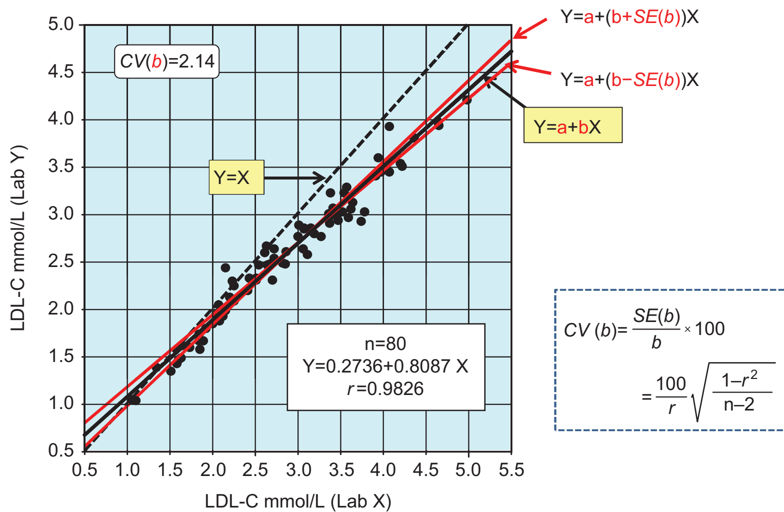
Test item	Laboratory A	Laboratory B	Laboratory C	Laboratory D
General chemistry				
TP	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
Alb	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
UN	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
UA	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
CRE	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
TBil	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
Na, K, Cl	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
Ca	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
IP	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
Fe	Abbott Architect C16000		BC AU680	Roche Modular
Glu	Abbott Architect C16000	Roche Modular		Roche Modular
TG	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
TCho	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
HDL-C	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
LDL-C	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
AST	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
ALT	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
LD	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
ALP	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
GGT	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
CK	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
AMY	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
Immunoturbidimetry				
C3		Roche Modular	BC AU680	BC AU680
C4		Roche Modular	BC AU680	BC AU680
Tf	Abbott Architect C16000	Roche Modular	BC AU680	Roche Modular
TTR	Abbott Architect C16000	Roche Modular	BC AU680	
Immunoassay				
Ferritin	Abbott Architect i2000	Siemens Centaur	BC DxI800	BC DxI800
Folate	Abbott Architect i2000		BC DxI800	BC DxI800
VB12	Abbott Architect i2000		BC DxI800	BC DxI800
AFP	Abbott Architect i2000	BC DxI800	BC DxI800	
CEA	Abbott Architect i2000	Siemens Centaur	BC DxI800	BC DxI800
CA19-9	Abbott Architect i2000		BC DxI800	BC DxI800
CA125	Abbott Architect i2000	Roche E170	BC DxI800	
PSA	Abbott Architect i2000	Roche E170	BC DxI800	Roche Modular
Estradiol	Abbott Architect i2000	BC DxI800	BC DxI800	Roche Modular
PRL	Abbott Architect i2000	Siemens Centaur	BC DxI800	
LH	Abbott Architect i2000	Roche Modular	BC DxI800	BC DxI800
FSH	Abbott Architect i2000	Roche E170	BC DxI800	Roche Modular
TSH	Abbott Architect i2000	Roche E170	BC DxI800	Roche Modular
FT4	Abbott Architect i2000	Roche E170	BC DxI800	Roche Modular
FT3	Abbott Architect i2000		BC DxI800	Roche Modular
PTH	Abbott Architect i2000	Roche E170	BC DxI800	Roche Modular
Insulin		Siemens Centaur	BC DxI800	Roche Modular

Table 1 List of analytes and assay platforms for comparative measurements.

[2–6]; it was derived in this study, as shown in Appendix A. *SE*(*b*) can be standardized by taking its ratio to *b* and expressed as *CV*(*b*) (Figure 1)

$$SE(b)=\frac{b}{r}\sqrt{\frac{1-r^2}{\mathrm{n}-2}}$$
$$CV(b)=\frac{SE(b)}{b}\times 100=\frac{100}{r}\sqrt{\frac{1-r^2}{\mathrm{n}-2}}.$$

SE(*LL*) and *SE*(*UL*) can also be calculated theoretically using the formulae shown in Appendix B by specifying $x=LL_0$ and $x=UL_0$, respectively, together with well-known parameters of linear regression analysis: *r*, n, SD of the observed *x* (s_x), and SD around regression line computed by the ordinary least-squares method (s_{LS}, often denoted as $s_{y,x}$). The computed values of the precision parameters *CV*(*b*), *SE*(*LL*), and *SE*(*UL*) were compared with those obtained empirically by the conventional method using the bootstrap method



$$CV(b)=\frac{SE(b)}{b}\times 100 = \frac{100}{r}\sqrt{\frac{1-r^2}{n-2}}$$

Figure 1 The concept for *CV*(*b*) as a measure of precision of regression line.
In estimating the regression line using reduced MAR, the SE of *b* is given as a function of the correlation coefficient *r* and data size n. *CV*(*b*) can be a marker in predicting the precision of converting the RI. In the graph, the range of fluctuation of the regression line according to $b\pm SE(b)$ is shown by the red line.

(8). In the computation, LL_0 and UL_0 were derived from the test results (n=80) of laboratory X as $\bar{x}\ \pm 1.96s_x$.

To standardize the variability of $SE(LL)$ or $SE(UL)$, we considered two methods. One is to take its ratio to the SD comprising RI (SD_{RI}), and we designated it as the SD ratio (SDR) of *LL* or *UL* (SDR_{LL} and SDR_{UL}, respectively) (Figure 2).

$$SDR_{LL}=\frac{SE(LL)}{SD_{RI}}\times 100,\ SDR_{UL}=\frac{SE(UL)}{SD_{RI}}\times 100\left(SD_{RI}=\frac{UL-LL}{3.92}\right),$$

where SD_{RI} can be obtained as the width of the RI after conversion, $UL-LL$, divided by 3.92. SD_{RI} corresponds to $\sqrt{SD_I^2+SD_G^2}$, to specify the level of analytical bias (B_A) in clinical chemistry. SD_I and SD_G denote within- and between-individual SDs of test results. The following thresholds of B_A specified by Fraser [9] are commonly used in clinical chemistry: $B_A \le 0.25\sqrt{SD_I^2+SD_G^2}$ is regarded as desirable performance, and $B_A \le 0.125\ \sqrt{SD_I^2+SD_G^2}$ as optimal performance. In analogy, we regarded SDR ≤ 0.25 as the level of precision allowable for converting the RI, and SDR ≤ 0.125 as optimal precision.

To avoid the influence of outlying points in the regression, the SD of data points along the minor axis of the ellipse (SD_{Min}) perpendicular to the regression line was computed, and any point four times SD_{Min} or more away from the center along the minor axis was excluded from the computation.

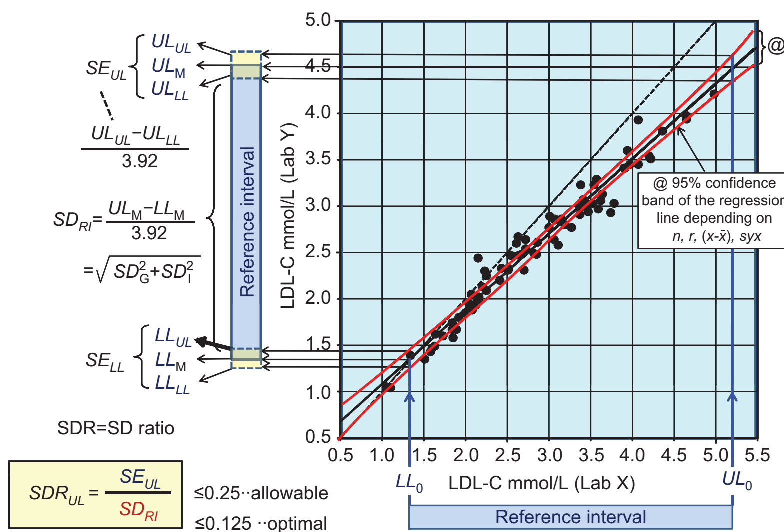


Figure 2 The concept of SDR as a measure of precision for converting the RI.
The RI determined by laborator X (LL_0-UL_0) is converted to that of laboratory Y as $LL-UL$ using the reduced major-axis regression line. The SEs of *LL* and *UL* [*SE*(*LL*) and *SE*(*UL*), respectively] can be estimated by the theoretical formulae. *SE*(*LL*) and *SE*(*UL*) are standardized as a ratio to the SD comprising the RI (SD_{RI}). We call the ratios SDR_{LL} and SDR_{UL}.

When the distribution of observed test results deviates from a Gaussian distribution, the theoretically derived $CV(b)$, SDR_{LL}, and SDR_{UL} can be biased partly due to heteroscedasticity of data points around the regression line (more scattering of data points toward the tail of the distribution). To examine the effect of Gaussian transformation on the regression statistics, we chose the following analytes whose values usually show a distribution with a long right tail: TBil, Glu, TG, AST, ALT, LD, GGT, CK, Tf, VB12, AFP, CEA, CA19-9, CA125, PSA, estradiol, PRL, LH, FSH, TSH, FT4, and insulin.

SDR_{LL} or SDR_{UL} based on the regression line derived after logarithmic (log) transformation can be computed from LL, UL, $SE(LL)$, and $SE(UL)$ using the following formulae involving exponential function for back-transformation, where LL^T and UL^T represent LL_0 and UL_0 converted using the regression line and $\bar{y}$ and $\bar{x}$ represent the averages of log(x) and log(y) for all data points.

$$LL^T=\bar{y}+b\times\left(\log\left(LL_0\right)-\bar{x}\right),\quad UL^T=\bar{y}+b\times\left(\log\left(UL_0\right)-\bar{x}\right)$$

$$SE\left(LL\right)=\frac{\exp\left(LL^T+1.96\times SE\left(LL^T\right)\right)-\exp\left(LL^T-1.96\times SE\left(LL^T\right)\right)}{3.92}$$

$$SE\left(UL\right)=\frac{\exp\left(UL^T+1.96\times SE\left(UL^T\right)\right)-\exp\left(UL^T-1.96\times SE\left(UL^T\right)\right)}{3.92}$$

$$LL=\exp(LL^T)\quad UL=\exp(UL^T)$$

$$SDR_{LL}=\frac{SE\left(LL\right)}{\frac{UL-LL}{3.92}}\quad SDR_{UL}=\frac{SE\left(UL\right)}{\frac{UL-LL}{3.92}}.$$

When a general power transformation (x^p) is used to convert the distribution of test values into the Gaussian form, the above formulae should be modified at the back-transformation step using the value to the power of $1/p$.

Resampling studies for evaluating the minimum sample size required for the panel

For each of the pairwise comparisons among the four laboratories, consisting of 80 pairs of test results, a subset of the results with data size of 20, 30, 40, 50, 60, or 70 was randomly sampled with replacement, and $CV(b)$, SDR_{LL}, and SDR_{UL} were computed by reduced MAR. In any case, the LL_0 and UL_0 required for computing LL and UL were fixed to the mean±1.96 SD derived from all test results (n=80) of laboratory X.

Results

Analytical variations

The analytical CV (CV_A) listed in Supplementary Table 1, which accompanies the article at http://www.degruyter.com/view/j/cclm.2013.51.issue-5/issue-files/cclm.2013.51.issue-5.xml, was computed for each analyte from the results of repeated measurements of the same panel done in each of the four laboratories. It represents the proportion of an average SD of 80 data points to the grand mean (GM) of all observations, which corresponds to coarse between-day CV of test results, including the component of within-day CV. The observed magnitude of CV_A for almost all analytes was well below 0.75 (Δ), 0.5 (○), or 0.25(●) times the within-individual CV (CV_I), as reported by Fraser [9], regarded as the thresholds of minimal, desirable, and optimal levels of error, respectively. CV_A above the allowable limit (>0.75×CV_I) was found for Na, Cl, Ca, estradiol, FSH, and FT4 in some laboratories.

The assay precision of each analyte was also analyzed from the test results of quality control sera measured concurrently over the study period. The results are summarized in Supplementary Table 2. Among the standardized analytes, we regarded Na, Cl, and Ca as difficult analytes for keeping the precision within acceptable levels. As for the non-standardized analytes (mostly immunoassays), fairly large between-day variations were observed for ferritin, CA19-9, and FT4. However, for other analytes, we regarded the reproducibility as excellent, considering the inherent variability of immunoassays, with their CV_A well below CV_I×0.5.

Comparison of test results across the four laboratories

Typical examples of the comparative measurements of the panel (n=80) among the four laboratories are shown in Figure 3 for two analytes, HDL-C and TSH. The full results are available in Supplementary Figure 1 composed of 45 correlation matrix (8×8) graphs. Because each laboratory (A, B, C, and D) measured the panel twice 1 week apart, the number 1 or 2 attached to the laboratory name indicates the first or second measurement. The regression lines (depicted in solid lines) were computed by reduced MAR. The broken lines represent the line of concordance ($y=x$). The colored blocks of the matrix indicated reproducibility of test results within the same laboratory.
The following findings can be noted from Supplementary Figure 1:

- The reproducibility of two measurements obtained over a 1-week period within the same laboratory is excellent by visual inspection except for those of Na, Cl, CA19-9, FT4, and FT3 for which a large scatter around the regression lines was observed.
- The regression lines drawn between the results from any two laboratories were very close to $y=x$ for almost all of the standardized analytes except for those showing variable degrees of bias, judging from a wide range of slopes (cf. Table 2), such as Na, Cl, Ca, HDL-C, LDL-C, ALP, and C4.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Test item	Correlation coefficient *r*	90% CI of slope *b*	*CV*(*b*)	SDR_{LL}	SDR_{UL}	AveSDR	SDR_{UL}/SDR_{LL}	Skewness	bCV(*b*)	bSDR_{LL}	bSDR_{UL}	bCV(*b*)/*CV*(*b*)	bSDR_{LL}/SDR_{LL}	bSDR_{UL}/SDR_{UL}
TP	0.932	0.919–1.088	4.43	0.097	0.096	0.096	0.99	-0.137	4.795	0.101	0.106	1.083	1.046	1.112
Alb	0.897	0.831–1.203	5.57	0.121	0.121	0.121	1.00	0.009	4.573	0.101	0.106	0.821	0.836	0.882
UN	0.985	0.967–1.034	2.01	0.043	0.047	0.045	1.08	0.367	2.029	0.040	0.049	1.008	0.921	1.049
UA	0.991	0.941–1.063	1.46	0.031	0.032	0.031	1.03	0.041	1.324	0.029	0.031	0.910	0.922	0.977
CRE	0.963	0.731–1.367	3.29	0.065	0.074	0.069	1.14	0.149	3.028	0.062	0.065	0.920	0.958	0.883
TBil	0.987	0.874–1.144	1.86	0.037	0.047	0.042	1.26	0.742	2.035	0.036	0.055	1.092	0.963	1.168
Na	0.485	0.450–2.230	20.41	0.411	0.425	0.418	1.03	0.095	8.970	0.195	0.221	0.440	0.475	0.520
K	0.982	0.931–1.075	2.21	0.049	0.048	0.049	1.00	0.240	1.848	0.039	0.043	0.835	0.804	0.891
Cl	0.758	0.537–1.863	9.76	0.207	0.210	0.208	1.02	0.039	7.467	0.156	0.163	0.765	0.756	0.775
Ca	0.897	0.849–1.178	5.61	0.121	0.121	0.121	1.00	−0.082	4.668	0.096	0.109	0.832	0.791	0.898
IP	0.973	0.942–1.062	2.65	0.058	0.058	0.058	1.00	−0.062	2.303	0.045	0.048	0.868	0.781	0.828
Fe[a]	0.997	0.932–1.073	0.86	0.018	0.020	0.019	1.06	0.380	0.893	0.019	0.020	1.042	1.014	1.021
Glu[a]	0.993	0.949–1.054	1.34	0.028	0.033	0.030	1.18	1.020	1.701	0.036	0.043	1.269	1.283	1.317
TCho	0.994	0.953–1.049	1.26	0.028	0.028	0.028	1.00	−0.106	1.174	0.023	0.029	0.930	0.847	1.044
TG	0.998	0.943–1.061	0.64	0.012	0.018	0.015	1.46	0.915	1.024	0.015	0.029	1.595	1.202	1.636
HDL-C	0.989	0.788–1.269	1.69	0.035	0.040	0.038	1.16	0.464	1.741	0.032	0.045	1.032	0.920	1.102
LDL-C	0.988	0.794–1.259	1.70	0.037	0.037	0.037	1.00	0.002	1.345	0.026	0.034	0.789	0.704	0.919
AST	0.987	0.934–1.070	1.86	0.036	0.052	0.044	1.45	1.347	2.111	0.037	0.061	1.136	1.026	1.173
ALT	0.996	0.828–1.208	0.96	0.018	0.028	0.023	1.59	0.967	1.180	0.019	0.036	1.232	1.074	1.281
ALP	0.974	0.260–3.841	2.68	0.058	0.061	0.059	1.05	0.309	3.732	0.062	0.097	1.394	1.059	1.598
LDH	0.961	0.840–1.191	3.00	0.065	0.066	0.066	1.01	0.607	2.075	0.038	0.051	0.692	0.582	0.782
GGT	0.997	0.886–1.129	0.82	0.015	0.023	0.019	1.52	0.687	1.067	0.015	0.033	1.301	0.951	1.417
CK	0.999	0.906–1.104	0.49	0.009	0.015	0.012	1.67	1.664	1.049	0.011	0.034	2.141	1.210	2.288
AMY	0.992	0.893–1.120	1.46	0.032	0.032	0.032	1.00	0.347	1.333	0.022	0.034	0.916	0.680	1.053
C3[a]	0.955	0.918–1.090	3.50	0.075	0.078	0.077	1.04	0.229	3.177	0.067	0.074	0.907	0.887	0.939
C4[a]	0.972	0.801–1.248	2.73	0.059	0.061	0.060	1.04	0.513	2.497	0.046	0.065	0.916	0.788	1.068
Tf	0.982	0.831–1.203	2.16	0.046	0.048	0.047	1.04	0.403	2.295	0.045	0.053	1.062	0.969	1.098
TTR[a]	0.977	0.926–1.079	2.58	0.055	0.059	0.057	1.07	0.225	2.653	0.045	0.065	1.027	0.813	1.108
Ferritin	0.992	0.741–1.350	1.44	0.027	0.041	0.034	1.51	0.668	2.189	0.021	0.068	1.519	0.785	1.668
Folate[a]	0.976	0.704–1.421	2.56	0.054	0.059	0.057	1.09	0.307	2.797	0.045	0.072	1.092	0.824	1.228
VB12[a]	0.980	0.674–1.485	2.29	0.047	0.055	0.051	1.16	0.460	2.397	0.037	0.066	1.045	0.778	1.192
AFP[a]	0.967	0.908–1.102	3.00	0.060	0.075	0.067	1.24	0.610	3.205	0.050	0.086	1.067	0.836	1.147
CEA	0.922	0.970–1.030	4.78	0.094	0.120	0.107	1.28	0.416	5.151	0.066	0.146	1.077	0.704	1.217
CA19-9[a]	0.745	0.541–1.850	10.20	0.172	0.304	0.238	1.77	0.977	15.009	0.124	0.470	1.472	0.724	1.549
CA125[a]	0.927	0.905–1.105	4.62	0.091	0.118	0.105	1.29	0.705	5.202	0.070	0.149	1.126	0.762	1.257
PSA	0.996	0.931–1.074	1.22	0.024	0.036	0.030	1.46	0.941	1.939	0.019	0.060	1.585	0.780	1.683
Estradiol	0.987	0.736–1.359	1.87	0.034	0.062	0.048	1.84	1.236	2.334	0.031	0.080	1.245	0.920	1.288
PRL	0.935	0.610–1.639	3.91	0.068	0.118	0.093	1.74	1.429	4.899	0.042	0.155	1.254	0.616	1.311
LH	0.994	0.662–1.510	1.29	0.023	0.040	0.032	1.73	1.068	2.300	0.019	0.080	1.783	0.832	1.966

(Table 2 Continued)

(1) Test item	(2) Correlation coefficient *r*	(3) 90% CI of slope *b*	(4) *CV(b)*	(5) SDR_{LL}	(6) SDR_{UL}	(7) AveSDR	(8) SDR_{UL}/SDR_{LL}	(9) Skewness	(10) b*CV(b)*	(11) bSDR_{LL}	(12) bSDR_{UL}	(13) b*CV(b)*/*CV(b)*	(14) bSDR_{LL}/SDR_{LL}	(15) bSDR_{UL}/SDR_{UL}
FSH	0.997	0.597–1.679	0.86	0.015	0.029	0.022	1.88	0.920	1.522	0.007	0.056	1.764	0.433	1.944
TSH	0.979	0.813–1.230	2.40	0.050	0.060	0.055	1.20	0.767	3.086	0.043	0.083	1.286	0.861	1.394
FT4	0.779	0.746–1.341	9.13	0.192	0.198	0.195	1.03	0.462	7.809	0.148	0.185	0.855	0.770	0.935
FT3[a]	0.722	0.882–1.133	10.84	0.228	0.230	0.229	1.01	0.097	7.793	0.159	0.177	0.719	0.698	0.769
PTH	0.975	0.589–1.698	2.59	0.054	0.062	0.058	1.15	0.239	2.889	0.043	0.080	1.117	0.802	1.288
Insulin[a]	0.968	0.738–1.354	2.96	0.054	0.086	0.070	1.59	1.137	4.843	0.045	0.148	1.637	0.837	1.717

Table 2 Summary of regression line statistics for all pairwise comparisons of test results.
All sera in the panel (n=80) were measured twice by the four laboratories (or by three laboratories for test items marked by superscript a), and all pairwise comparisons were made excluding within-laboratory comparison. Therefore, the total number of comparisons for each analyte was either 24 or 12. Columns 2, 4–6, and 10–12: The statistics with gray headers are medians of those obtained by the all pairwise comparisons. Column 4: The CV of the regression line slope, *CV(b)*. Values of 5.5%–11% and >11% are marked by light and dark gray, respectively. Columns 5–7: The SDR_{LL} and SDR_{UL} represent the ratios of SD of the lower and upper limits (*LL* and *UL*) to the SD comprising RI, or (*UL*–*LL*)/3.92. The aveSDR represents the average of SDR_{LL} and SDR_{UL}. Values of 0.125–0.25 and >0.25 are marked by light and dark gray, respectively. Column 8: The ratio of SDR_{UL} to SDR_{LL} above 1.2 is marked. Column 9: Skewness (*Sk*) was determined from the distribution of test values measured by each laboratory. The average of *Sk* of the four laboratories is shown. Graded gray colors indicate $0.4<Sk\leq0.8$ and $0.8<Sk$. Columns 10–12: *CV(b)*, SDR_{LL}, and SDR_{UL} were determined by the bootstrap method by resampling 80 points for 500 times. Columns 13–15: The ratios of statistics obtained by the bootstrap method to those computed by the theoretical formulae. Values ≤0.8 or ≥1.2 are marked by gray background.

- For the non-standardized analytes, regression were approximately $y=x$ for IP, Fe, C3, AFP, CEA, CA125, and PSA, whereas variable degrees of bias were noted for ferritin, folate, VB12, CA19-9, estradiol, PRL, LH, FSH, FT4, PTH, and insulin.
- Although most points are close to the regression line, a few to several outlying points were observed in the plots for CEA, CA125, folate, PRL, and insulin.

Comparison of regression parameters representing the precision of converting RI

Table 2 shows the summary statistics obtained from the dataset shown in Figure 3 (and Supplementary Figure 1), which represent the closeness of fit of data to the regression lines and precision of conversion of the RIs when all test results (n=80) were used for the comparison. For each comparison between laboratory X and laboratory Y (X, Y=A, B, C, or D), statistics *r*, *b*, *CV*(*b*), SDR_{LL}, and SDR_{UL} were computed.

The total number of pairwise comparisons made (excluding within-laboratory comparison) was 24 [=[8×(8–1)–4×2]/2] for the analytes measured by all four laboratories and 12[=[6×(6–1)–3×2]/2] for those measured only by three laboratories. The statistics (listed in columns with gray headers numbered 2, 4–6, and 10–12) are the medians of those obtained by all comparisons.

The correlation coefficients *r* were generally very high ($r>0.95$). Analytes with $r<0.9$ include Alb, Na, Cl, Ca, CA19-9, FT4, and FT3, indicated by a gray background. The wider 90% range of variation of the slope *b* shown in column 3 implies either a lack of close correlation or the presence of between-method bias in the test results or both. The presence of both conditions implies that the conversion of test results between laboratories is not feasible.

The skewness (*Sk*) of distribution was computed from the test values of each laboratory, and the average *Sk* of the 4 (or 3) laboratories is shown in column 3. The distribution was regarded as nearly symmetrical when $|Sk|\leq0.4$, and values of $Sk>0.4$ are shown with a gray background.

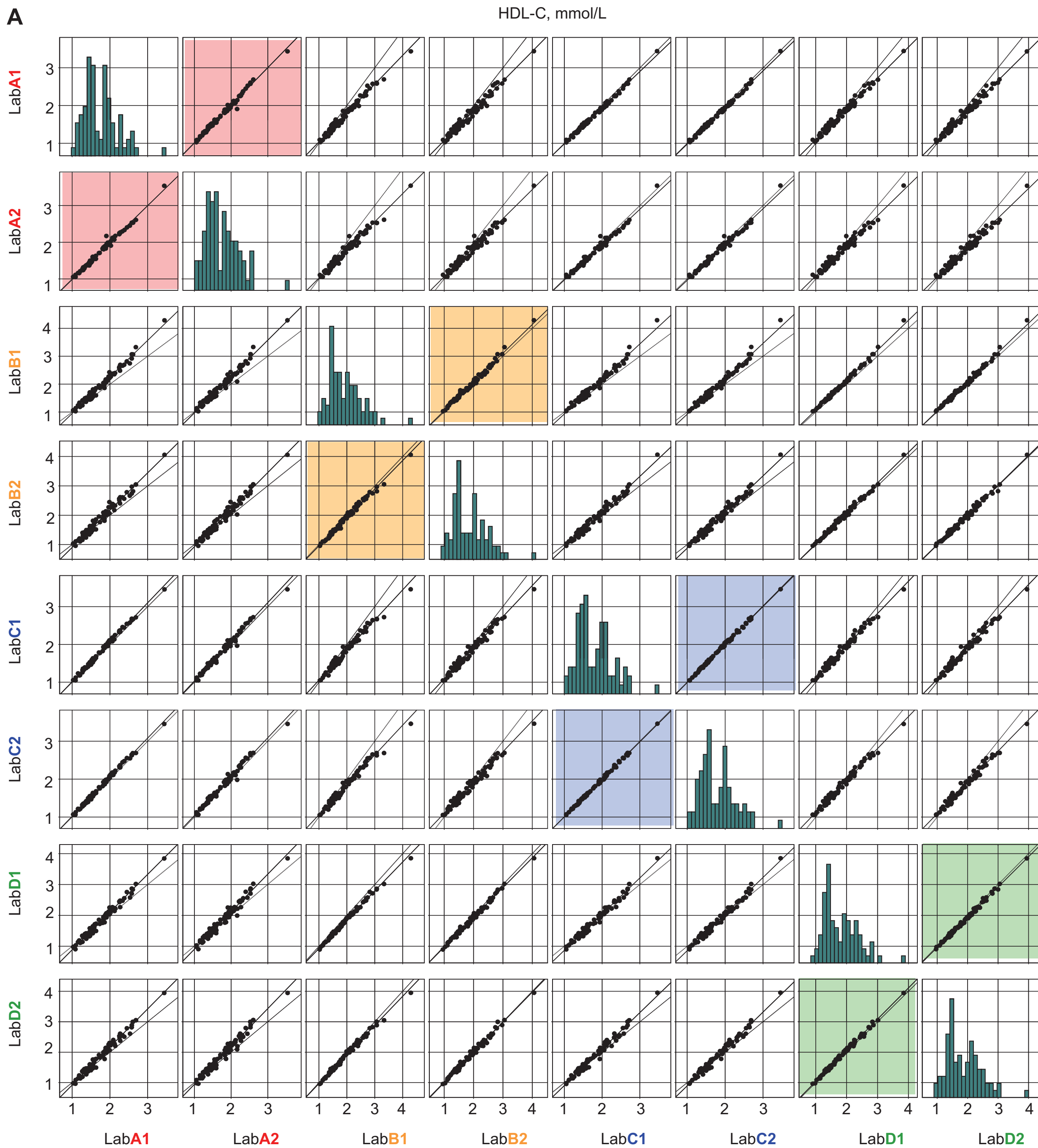
SDR_{LL} and SDR_{UL}, which were above the optimal (0.125) or allowable (0.25) limits, are also shown with a gray background. Analytes with either of the SDRs above the allowable limits using all test results (n=80) include Na and CA19-9; additional analytes with either of SDRs above the optimal limits are Cl, CRP, FT4, and FT3.

The ratio of SDR_{UL} to SDR_{LL} shown in column 8 was noticeably higher when *Sk* is >0.4. Close correlation ($r=0.874$) was observed between *Sk* and SDR_{UL}/SDR_{LL}, implying that SDRs depend on the distribution pattern.

Meanwhile, it was noted that $CV(b)$ shows a high correlation with both SDR_{LL} and SDR_{UL}. Furthermore, the correlation is dependent on Sk of the distribution of test results, as shown by analyzing the correlation after stratification of 45 analytes by Sk (Figure 4). It was also found that there is a closer correlation of $CV(b)$ with the average of SDR_{LL} and SDR_{UL} (aveSDR, shown in column 7), which was less affected by Sk than was the individual SDR. The relationship between the two statistics was computed as follows using reduced MAR:

$$\text{aveSDR}=0.001+0.0222\ CV(b)$$

From this equation, SDRs of 0.125 (optimal limit) and 0.25 (allowable limit) was calculated as equivalent to $CV(b)$ of approximately 5.5% and 11%, respectively.



(Figure 3 Continued)

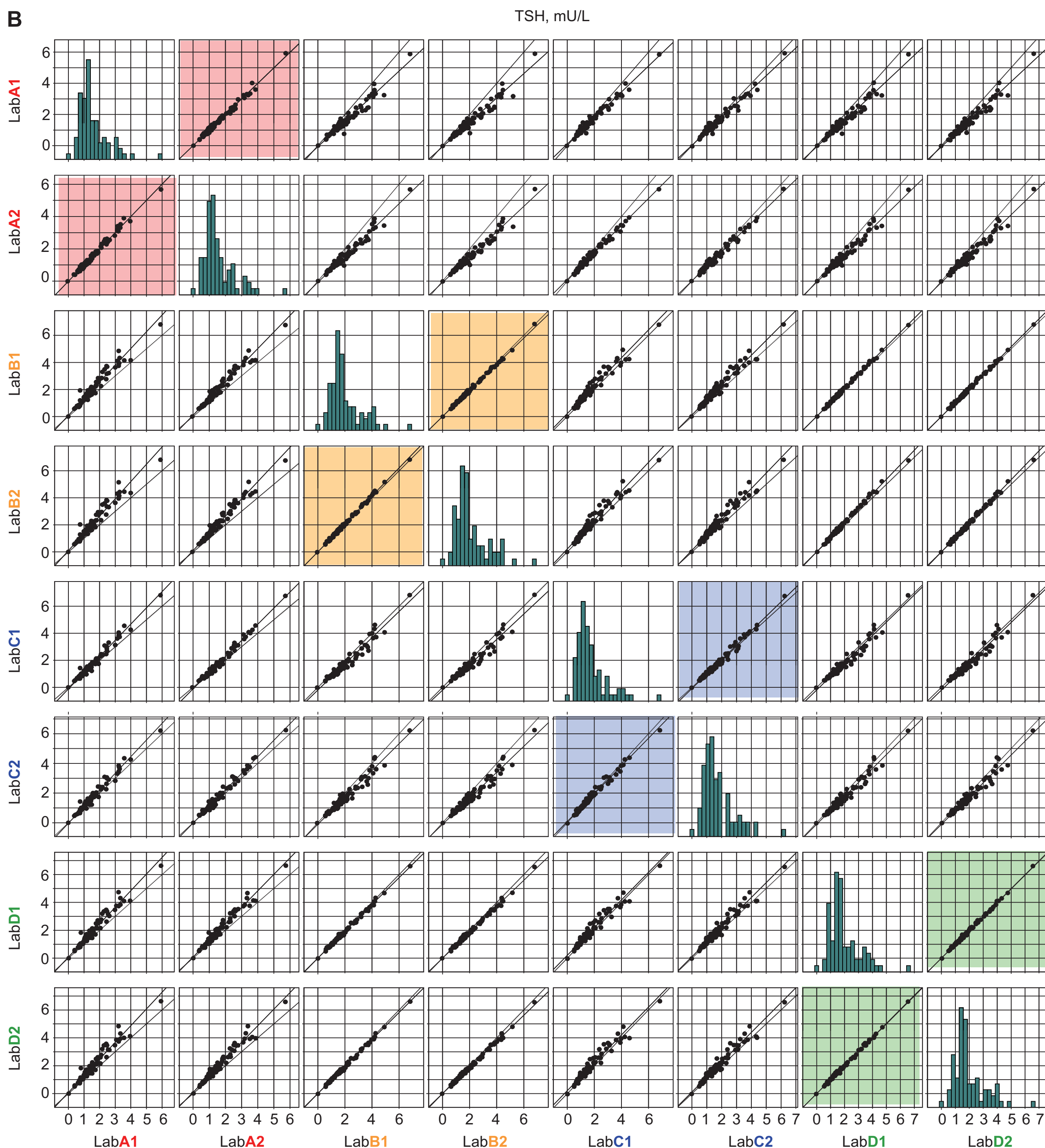


Figure 3 Examples of correlation matrix graphs of method comparison among the four laboratories.
Laboratories A, B, C, and D measured the panel of 80 sera twice, 1 week apart. The colored blocks indicate the within-laboratory variability of the test results. Within each graph, the line of concordance (y=x) is depicted by the broken line.

For validation of the theoretically computed values, $CV(b)$, SDR_{LL}, and SDR_{UL} were also estimated by the bootstrap method by resampling 80 data points 500 times. The results were designated as bCV(b), bSDR_{LL}, and bSDR_{UL} and are shown in columns 10–12.

Their ratios to respective values calculated by the formulae shown in Appendix A are listed in columns 13–15. It was found that bCV(b)/CV(b) and bSDR_{UL}/SDR_{UL} depend on Sk; when the distribution is not skewed (|Sk|<0.4), both ratios are nearly 1.0, indicating that the value derived by

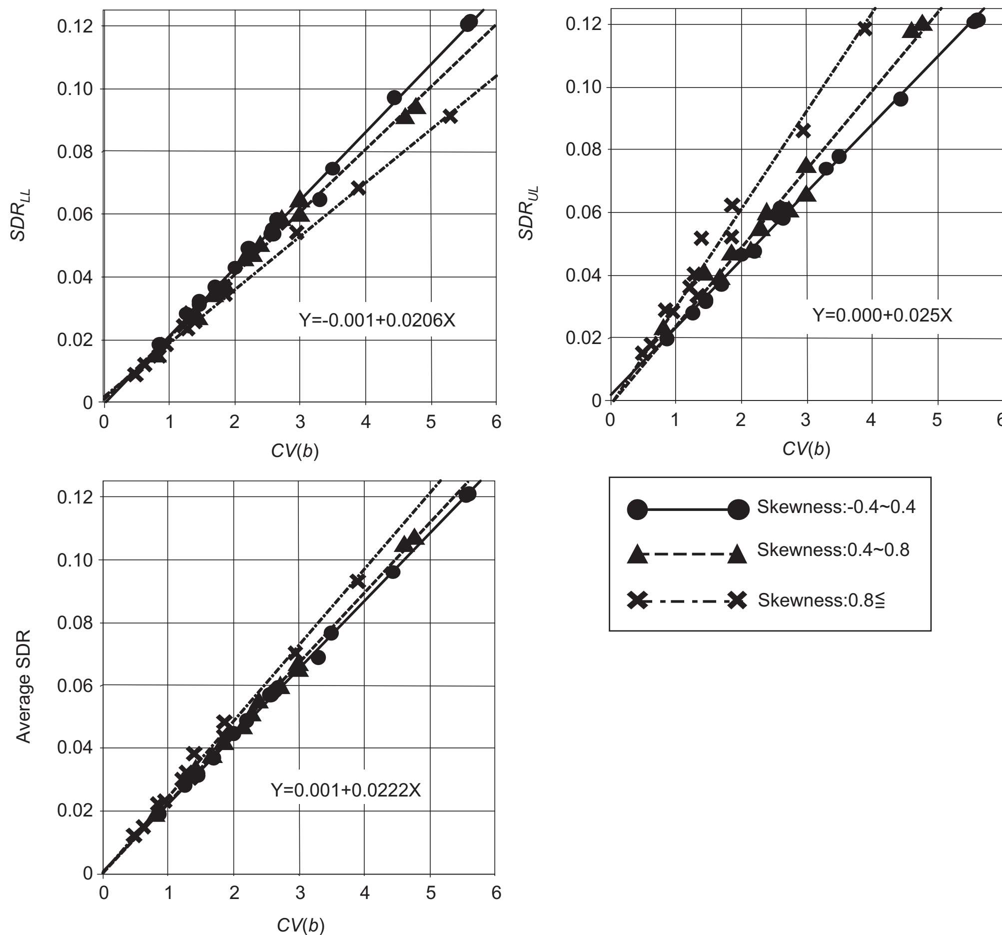


Figure 4 Relationships among SDRs and *CV*(*b*).
From the data listed in Table 2, the relationships between *CV*(*b*) and SDRs (SDR_{LL}, SDR_{UL}, or the average of the two) were examined after partitioning the data for 45 analytes by average skewness (*Sk*) of their distributions into three groups ($|Sk|\leq0.4$, $0.4<Sk\leq0.8$, and $0.8<Sk$). The regression lines drawn were computed by reduced MAR.

the bootstrap method matches well with the calculated value from the formula. When the distribution is skewed to the higher side ($Sk>0.4$), both ratios increase in proportion to *Sk*.

Meanwhile, bSDR_{LL}/SDR_{LL} is not much associated with *Sk*. This implies that skewness to the lower side of values does not affect the regression in estimating *LL* because of clustering of data to the lower side.

Effects of log transformation of the skewed distribution on the regression line statistics

Table 3 shows a summary of the regression line statistics obtained either with or without log transformation of results for the selected analytes. As evident from the comparison of columns 1 and 2, the skewness was greatly decreased after log transformation. *CV*(*b*) calculated by the formula did not change much as a whole, whereas bCV(*b*) empirically derived by the bootstrap method was generally decreased by the transformation, suggesting an overestimation of the variation of the slope. In contrast, SDR_{UL} increased after the transformation, whereas bSDR_{UL} changed very little. This phenomenon is interpreted as follows: in the former, by log transformation, the data points around *LL* have more influence on the regression, and thus, a wider variation occurred to the predicted *UL*. In the latter, a wider variation of data points around *UL* was reduced after log transformation, but at the same time, the variation around *LL* was increased, and thus, as a whole, the average bSDR_{UL} did not change much. SDR_{LL} is reduced after log transformation by either of the methods, indicating that both SDR_{LL} and bSDR_{LL} tend to be underestimated when the data are highly skewed. In any case, after log transformation of test results, as shown in columns 24 and 26, the dissociation of the two approaches in evaluating the precision of regression errors is greatly reduced.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)
	Skewness		*CV*(*b*)			SDR_{LL}			SDR_{UL}			b*CV*(*b*)			bSDR_{LL}			bSDR_{UL}			b*CV*(*b*)/ *CV*(*b*)		bSDR_{LL}/ SDR_{LL}		bSDR_{UL}/ SDR_{UL}	
Log transform	(–)	(+)	(–)	(+)	Ratio	(–)	(+)	Ratio	(–)	(+)	Ratio	(–)	(+)	Ratio	(–)	(+)	Ratio	(–)	(+)	Ratio	(–)	(+)	(–)	(+)	(–)	(+)
TBil	0.74	0.12	1.86	2.46	1.32	0.037	0.019	0.52	0.047	0.098	2.07	2.04	2.05	1.01	0.036	0.021	0.58	0.055	0.076	1.38	1.09	0.83	0.96	1.07	1.17	0.78
Glu	1.02	0.29	1.34	2.01	1.50	0.028	0.021	0.77	0.033	0.039	1.20	1.70	1.53	0.90	0.036	0.022	0.61	0.043	0.047	1.09	1.27	0.76	1.28	1.02	1.32	1.20
TG	0.92	0.31	0.64	0.87	1.35	0.012	0.005	0.42	0.018	0.041	2.27	1.02	0.73	0.71	0.015	0.006	0.40	0.029	0.027	0.92	1.59	0.84	1.20	1.15	1.64	0.67
AST	1.35	0.58	1.86	3.13	1.69	0.036	0.027	0.76	0.052	0.081	1.58	2.11	2.46	1.17	0.037	0.031	0.84	0.061	0.085	1.40	1.14	0.78	1.03	1.13	1.17	1.04
ALT	0.97	0.40	0.96	1.88	1.96	0.018	0.009	0.50	0.028	0.082	2.88	1.18	1.47	1.24	0.019	0.010	0.53	0.036	0.068	1.86	1.23	0.78	1.07	1.14	1.28	0.83
LDH	0.61	0.09	3.00	2.09	0.70	0.065	0.025	0.38	0.066	0.051	0.78	2.08	1.78	0.86	0.038	0.024	0.64	0.051	0.054	1.06	0.69	0.85	0.58	0.99	0.78	1.06
GGT	0.69	0.19	0.82	0.93	1.14	0.015	0.005	0.33	0.023	0.049	2.08	1.07	0.79	0.74	0.015	0.005	0.36	0.033	0.035	1.05	1.30	0.85	0.95	1.05	1.42	0.71
CK	1.66	0.54	0.49	0.92	1.88	0.009	0.005	0.53	0.015	0.030	2.00	1.05	0.73	0.69	0.011	0.005	0.48	0.034	0.033	0.97	2.14	0.79	1.21	1.10	2.29	1.10
Tf	0.40	0.16	2.16	1.96	0.91	0.046	0.030	0.64	0.048	0.048	0.99	2.30	1.67	0.73	0.045	0.028	0.63	0.053	0.046	0.86	1.06	0.85	0.97	0.96	1.10	0.95
VB12	0.46	-0.09	2.29	2.28	0.99	0.047	0.017	0.37	0.055	0.085	1.54	2.40	1.84	0.77	0.037	0.018	0.50	0.066	0.075	1.14	1.04	0.81	0.78	1.05	1.19	0.88
AFP	0.61	-0.10	3.00	2.46	0.82	0.060	0.018	0.29	0.075	0.107	1.43	3.21	2.21	0.69	0.050	0.019	0.38	0.086	0.100	1.16	1.07	0.90	0.84	1.08	1.15	0.93
CEA	0.42	-0.08	4.78	2.48	0.52	0.094	0.014	0.15	0.120	0.127	1.06	5.15	2.16	0.42	0.066	0.013	0.20	0.146	0.115	0.78	1.08	0.87	0.70	0.93	1.22	0.91
CA19-9	0.98	0.04	10.20	2.48	0.24	0.172	0.014	0.08	0.304	0.128	0.42	15.01	2.23	0.15	0.124	0.013	0.10	0.470	0.115	0.24	1.47	0.90	0.72	0.91	1.55	0.90
CA125	0.71	0.14	4.62	3.37	0.73	0.091	0.022	0.24	0.118	0.148	1.25	5.20	2.78	0.53	0.070	0.022	0.32	0.149	0.132	0.89	1.13	0.82	0.76	1.03	1.26	0.89
PSA	0.94	0.03	1.22	1.50	1.22	0.024	0.006	0.26	0.036	0.061	1.71	1.94	1.20	0.62	0.019	0.007	0.36	0.060	0.061	1.02	1.58	0.80	0.78	1.08	1.68	1.00
Estradiol	1.24	0.29	1.87	2.09	1.11	0.034	0.007	0.21	0.062	0.106	1.72	2.33	1.53	0.66	0.031	0.009	0.28	0.080	0.078	0.98	1.25	0.73	0.92	1.21	1.29	0.73
PRL	1.43	0.43	3.91	1.48	0.38	0.068	0.005	0.07	0.118	0.069	0.58	4.90	1.25	0.25	0.042	0.004	0.11	0.155	0.082	0.53	1.25	0.84	0.62	0.98	1.31	1.20
LH	1.07	0.30	1.29	0.95	0.74	0.023	0.002	0.07	0.040	0.076	1.87	2.30	0.87	0.38	0.019	0.002	0.09	0.080	0.072	0.91	1.78	0.91	0.83	1.01	1.97	0.95
FSH	0.92	0.46	0.86	0.63	0.73	0.015	0.001	0.04	0.029	0.058	2.01	1.52	0.52	0.34	0.007	0.001	0.09	0.056	0.049	0.87	1.76	0.83	0.43	0.98	1.94	0.84
TSH	0.77	0.08	2.40	2.07	0.86	0.050	0.012	0.24	0.060	0.102	1.71	3.09	1.72	0.56	0.043	0.011	0.25	0.083	0.102	1.23	1.29	0.83	0.86	0.89	1.39	1.00
FT4	0.46	0.11	9.13	8.22	0.90	0.192	0.120	0.63	0.198	0.204	1.03	7.81	5.68	0.73	0.148	0.095	0.65	0.185	0.166	0.90	0.86	0.69	0.77	0.79	0.93	0.81
Insulin	1.14	0.16	2.96	1.80	0.61	0.054	0.013	0.23	0.086	0.071	0.82	4.84	1.48	0.31	0.045	0.011	0.24	0.148	0.068	0.46	1.64	0.82	0.84	0.87	1.72	0.97
Averages					1.01			0.35			1.50			0.66			0.39			0.99	1.31	0.82	0.87	1.02	1.40	0.93

Table 3 The effect of logarithmic transformation on the regression statistics related to the precision of the method comparison.
Skewness (Sk) was determined from distribution of test values measured by each lab. The average of Sk of the 4 lab is shown. Graded gray colors indicate $0.4<Sk\le 0.8$ and $0.8<Sk$. *CV*(*b*) represents CV of the regression line slope computed by the theoretical formula. The median CV(b) of all pairwise comparison among 4 labs is shown. Values of 5.5–11%, and $>11\%$ are marked by light and dark gray. b*CV*(*b*) represents CV of the regression line slope empirically derived by the bootstrap method. The same scheme as *CV*(*b*) is used for listing of the values. SDR_{LL} and SDR_{UL} represent the ratio of *SE*(*LL*) and *SE*(*UL*) to the SD comprising RI, respectively, which were computed by the theoretical formula. The median SDRs of all pairwise comparison among 4 labs are shown. Values of 0.125–0.25 and >0.25 are marked by light and dark gray. bSDR_{LL} and bSDR_{UL} represent the ratio of *SE*(*LL*) and *SE*(*UL*) to the SD comprising RI, respectively, which were computed by the bootstrap method. The same scheme as SDR_{LL} and SDR_{UL} is used for listing of the values.

Resampling study to examine the effect of data size on precision of the conversion

From the original sets of 80 test results from laboratories X and Y, a sample subset (n=20, 30, 40, 50, 60, 70, or 80) was randomly drawn (resampled) with replacement, and the slope *b* by reduced MAR and converted LL_0 and UL_0 (*LL* and *UL*) were computed. This process was repeated 500 times, and *CV*(*b*), SDR_{LL}, and SDR_{UL} were computed using the theoretical formula.

Table 4 summarizes the results showing the effect of sample size on three statistics of conversion. It tabulates the medians SDR_{LL}, SDR_{UL}, and *CV*(*b*) of all pairwise comparisons for each data size (12 combination for analytes measured twice by three laboratories, and 24 combinations for those measured twice by four laboratories). From our criterion for an allowable level (SDR≤0.25), the percentage of analytes with that level of precision does not change much for 20≤n≤60. Meanwhile, the percentage of analytes with the optimum level of precision (SDR≤0.125) declines appreciably for n=20. For *CV*(*b*), applying the above-mentioned criteria [allowable *CV*(*b*) as ≤11%; optimum *CV*(*b*) as ≤5.5%], a similar tendency was observed to that for SDRs.

Assuming 40 sera as a practical size of the panel for use in the worldwide study, based on the allowable limits of *CV*(*b*), the reference values of Na, Cl, CA19-9, FT4, and FT3 (or 11% of the analytes examined) were not acceptable for use in comparison across the countries. At the optimal level of *CV*(*b*), 20% of the analytes, with the addition of TP, Alb, Ca, and CA125 to the above analytes, were regarded as not acceptable for comparison.

Discussion

The key strategy of the worldwide study on reference values is to measure a panel of sera in common and to align test results across countries on the basis of linear regression analysis. Thus, the objective of this preliminary investigation was to determine the most appropriate parameter along with its threshold value to judge its utility in converting RI among the collaborating laboratories. This study also intends to determine the most practical number of specimens (sample size, n) in the panel for use in the worldwide study.

There are several options for the statistical approach for the method comparison [2–7]. In the field of clinical chemistry, DR [7] has generally been considered as the most rational one because it can deal with the imbalance in the inherent errors of variables *x* and *y*. However, it requires an appropriate estimate of the error ratio (λ), especially when *x* and *y* variables are measured in different units or scales. Therefore, a replicate measurement is expected at the time of method comparison, which is not a common laboratory practice. Nevertheless, it is not reliable to estimate λ from the CV_A of each variable if not determined concurrently. Furthermore, λ is usually not constant over the range of regression. This situation necessitates the use of weighted DR [7] and proper derivation of the weighting factor, which further complicates the problem by requiring a larger number of measurements in replicate. Another downside of DR is that there is no mathematical formula to derive the SEs of the slope and other parameters. Their derivation depends on the use of the bootstrap method [8], which may not be reliable with a small sample size.

Meanwhile, the reduced MAR [2–7] gives the line of structural relationship equal to the reduced major axis of the ellipse delineating the scatter of data points. Its preferable property is that the regression is not influenced by the scale of measurement [2, 3] and thus can be used for comparison of test results for non-standardized analytes measured using two different units or scales. Furthermore, the SEs of the slope and point estimation based on the regression can be mathematically derived as we demonstrated in the appendices. Therefore, we regarded MAR as a practical method of choice for the purpose of the worldwide study.

The precision of the conversion using MAR depends on the sample size n and the degree of correlation of test results from the laboratories under comparison. Both parameters are contained in the mathematical formula to derive the CV of the slope: $CV(b)=100\sqrt{(1-r^2)/(n-2)}/r$.

CV(*b*), derived by the bootstrap method, was used in the 2009 Asian study to judge whether it was permissible to convert the centrally derived RI to that of each local laboratory [10, 11]. This conversion of RIs was provided as an option to the collaborating laboratories on an experimental basis. The number of sera actually measured for comparison in each local laboratory varied widely from nine to 73, with an average of 22.2. Therefore, we could not rely solely on the degree of correlation *r* to judge the appropriateness of converting the RI and found *CV*(*b*) to be useful in the judgment based on the knowledge that the CV of the slope by the least-squares method is well known to reflect both *r* and n. We empirically set the allowable level of conversion as *CV*(*b*)<10%, which roughly corresponds to *r*>0.96 (with n=10), *r* >0.90 (with n=25), or *r*>0.85 (with n=40), after inspection of all results of cross-check testing between the central laboratory and 44 local laboratories with regard to the stability of the regression lines.

Test item	SDR_{LL}*1							SDR_{UL}*1							CV(b)*2						
	Number of sera in the panel							Number of sera in the panel							Number of sera in the panel						
	20	30	40	50	60	70	80	20	30	40	50	60	70	80	20	30	40	50	60	70	80
TP	0.212	0.169	0.154	0.138	0.123	0.103	0.097	0.209	0.163	0.150	0.132	0.113	0.108	0.096	10.12	7.83	7.28	6.20	5.46	5.06	4.43
Alb	0.241	0.185	0.152	0.141	0.130	0.127	0.121	0.246	0.190	0.164	0.144	0.134	0.122	0.121	11.23	8.48	7.21	6.36	6.18	5.79	5.57
UN	0.080	0.071	0.059	0.055	0.048	0.046	0.043	0.089	0.072	0.064	0.058	0.050	0.049	0.047	3.76	3.22	2.65	2.58	2.20	2.15	2.01
UA	0.055	0.051	0.039	0.029	0.030	0.033	0.031	0.049	0.043	0.041	0.033	0.033	0.031	0.032	2.40	2.27	1.82	1.43	1.47	1.43	1.46
CRE	0.144	0.105	0.101	0.084	0.073	0.074	0.065	0.170	0.123	0.108	0.095	0.078	0.083	0.074	7.08	5.27	4.80	4.38	3.48	3.68	3.29
TBil	0.083	0.065	0.058	0.048	0.044	0.041	0.037	0.106	0.089	0.073	0.057	0.058	0.050	0.047	4.30	3.42	2.95	2.32	2.34	2.01	1.86
Na	0.713	0.570	0.544	0.473	0.444	0.476	0.411	0.777	0.631	0.585	0.542	0.464	0.515	0.425	35.30	29.72	27.14	24.55	23.28	24.99	20.41
K	0.090	0.077	0.066	0.060	0.056	0.054	0.049	0.087	0.076	0.066	0.059	0.054	0.053	0.048	4.02	3.54	3.12	2.69	2.49	2.42	2.21
Cl	0.437	0.342	0.325	0.266	0.233	0.213	0.207	0.450	0.353	0.320	0.277	0.243	0.226	0.210	20.65	16.51	15.64	12.53	10.98	10.12	9.76
Ca	0.246	0.201	0.162	0.156	0.131	0.126	0.121	0.240	0.201	0.174	0.160	0.140	0.128	0.121	11.18	9.45	7.81	7.48	6.16	5.81	5.61
IP	0.099	0.095	0.076	0.062	0.060	0.056	0.058	0.093	0.087	0.069	0.065	0.059	0.054	0.058	4.36	4.09	3.29	2.85	2.71	2.45	2.65
Fe	0.027	0.027	0.025	0.020	0.019	0.018	0.018	0.030	0.026	0.028	0.020	0.019	0.019	0.020	1.22	1.18	1.27	0.91	0.84	0.83	0.86
Glu	0.067	0.051	0.040	0.037	0.033	0.030	0.028	0.077	0.068	0.050	0.046	0.038	0.038	0.033	3.23	2.87	1.99	1.88	1.59	1.56	1.34
TCho	0.052	0.047	0.038	0.034	0.031	0.028	0.028	0.049	0.046	0.038	0.033	0.030	0.028	0.028	2.30	2.10	1.78	1.54	1.40	1.32	1.26
TG	0.026	0.022	0.018	0.016	0.015	0.012	0.012	0.041	0.032	0.028	0.026	0.023	0.019	0.018	1.40	1.14	1.02	0.91	0.81	0.67	0.64
HDL-C	0.064	0.057	0.051	0.044	0.040	0.036	0.035	0.077	0.069	0.057	0.048	0.044	0.041	0.040	3.24	2.88	2.40	2.02	1.88	1.72	1.69
LDL-C	0.062	0.050	0.048	0.041	0.034	0.035	0.037	0.063	0.053	0.047	0.039	0.037	0.038	0.037	2.91	2.37	2.19	1.94	1.64	1.66	1.70
AST	0.082	0.067	0.056	0.047	0.040	0.038	0.036	0.136	0.104	0.076	0.072	0.059	0.060	0.052	4.61	3.78	2.77	2.64	2.14	2.13	1.86
ALT	0.039	0.029	0.026	0.022	0.021	0.019	0.018	0.061	0.050	0.042	0.039	0.034	0.032	0.028	1.99	1.66	1.43	1.33	1.16	1.05	0.96
ALP	0.100	0.090	0.084	0.066	0.064	0.062	0.058	0.103	0.088	0.084	0.065	0.066	0.058	0.061	4.29	4.03	3.86	2.96	2.98	2.74	2.68
LDH	0.099	0.068	0.059	0.052	0.051	0.049	0.065	0.092	0.075	0.068	0.053	0.051	0.049	0.066	4.41	3.29	2.97	2.49	2.27	2.23	3.00
GGT	0.028	0.022	0.020	0.018	0.017	0.016	0.015	0.048	0.039	0.031	0.030	0.026	0.024	0.023	1.64	1.30	1.05	1.06	0.90	0.81	0.82
CK	0.016	0.015	0.012	0.012	0.010	0.010	0.009	0.030	0.028	0.025	0.024	0.017	0.016	0.015	1.01	0.92	0.85	0.81	0.55	0.54	0.49
AMY	0.043	0.038	0.035	0.031	0.029	0.027	0.032	0.044	0.038	0.034	0.032	0.029	0.026	0.032	2.01	1.72	1.57	1.42	1.31	1.20	1.46
C3	0.118	0.100	0.099	0.093	0.074	0.078	0.075	0.144	0.107	0.104	0.100	0.080	0.083	0.078	6.23	4.44	4.72	4.34	3.67	3.62	3.50
C4	0.082	0.070	0.072	0.070	0.055	0.059	0.059	0.085	0.081	0.069	0.076	0.053	0.060	0.061	3.78	3.60	3.18	3.32	2.38	2.65	2.73
Tf	0.086	0.070	0.061	0.056	0.050	0.046	0.046	0.089	0.071	0.063	0.056	0.050	0.047	0.048	4.02	3.36	2.80	2.56	2.30	2.12	2.16
TTR	0.111	0.091	0.069	0.065	0.061	0.058	0.055	0.120	0.097	0.077	0.074	0.069	0.067	0.059	5.52	4.35	3.26	3.21	2.94	2.89	2.58
Ferritin	0.043	0.035	0.030	0.027	0.028	0.025	0.027	0.061	0.051	0.047	0.042	0.045	0.038	0.041	2.16	1.81	1.68	1.49	1.57	1.37	1.44
Folate	0.084	0.093	0.080	0.069	0.061	0.060	0.054	0.101	0.107	0.085	0.075	0.067	0.062	0.059	4.23	4.74	3.75	3.25	2.95	2.80	2.56
VB12	0.086	0.069	0.066	0.057	0.054	0.050	0.047	0.108	0.086	0.075	0.071	0.064	0.059	0.055	4.45	3.56	3.18	2.92	2.74	2.44	2.29
AFP	0.132	0.091	0.085	0.073	0.066	0.060	0.060	0.157	0.118	0.101	0.095	0.081	0.076	0.075	6.12	4.78	4.22	3.69	3.28	2.99	3.00
CEA	0.119	0.105	0.103	0.099	0.087	0.087	0.094	0.157	0.138	0.138	0.123	0.111	0.110	0.120	6.22	5.52	5.41	5.01	4.37	4.46	4.78
CA19-9	0.236	0.246	0.216	0.210	0.177	0.169	0.172	0.535	0.429	0.384	0.334	0.323	0.322	0.304	18.05	14.43	12.83	12.44	11.60	11.30	10.20
CA125	0.150	0.123	0.109	0.094	0.101	0.084	0.091	0.213	0.159	0.142	0.133	0.134	0.115	0.118	8.68	6.27	5.62	5.20	5.30	4.53	4.62
PSA	0.032	0.026	0.024	0.024	0.021	0.021	0.024	0.058	0.040	0.036	0.034	0.029	0.030	0.036	2.14	1.33	1.31	1.19	1.03	1.06	1.22
Estradiol	0.066	0.052	0.044	0.040	0.038	0.033	0.034	0.136	0.104	0.083	0.077	0.069	0.067	0.062	4.02	3.05	2.44	2.18	2.06	1.95	1.87
PRL	0.051	0.050	0.051	0.040	0.033	0.030	0.068	0.095	0.095	0.088	0.075	0.075	0.058	0.118	3.18	3.14	2.90	2.45	2.39	1.88	3.91
LH	0.039	0.034	0.030	0.025	0.027	0.024	0.023	0.066	0.056	0.047	0.045	0.045	0.042	0.040	2.16	1.78	1.46	1.45	1.43	1.36	1.29

(Table 4 Continued)

Test item	SDR_{LL} *1							SDR_{UL} *1							$CV(b)$ *2						
	Number of sera in the panel							Number of sera in the panel							Number of sera in the panel						
	20	30	40	50	60	70	80	20	30	40	50	60	70	80	20	30	40	50	60	70	80
FSH	0.021	0.023	0.022	0.018	0.015	0.016	0.015	0.047	0.041	0.036	0.033	0.029	0.030	0.029	1.36	1.23	1.06	1.02	0.88	0.90	0.86
TSH	0.088	0.071	0.061	0.056	0.052	0.047	0.050	0.099	0.092	0.073	0.072	0.068	0.056	0.060	4.08	3.55	3.06	2.90	2.74	2.24	2.40
FT4	0.344	0.332	0.309	0.247	0.236	0.213	0.192	0.383	0.346	0.312	0.252	0.258	0.219	0.198	17.30	15.95	14.69	11.93	11.59	10.20	9.13
FT3	0.416	0.352	0.333	0.296	0.273	0.244	0.228	0.481	0.348	0.335	0.282	0.261	0.226	0.230	21.83	16.63	16.23	13.69	12.73	11.11	10.84
PTH	0.106	0.085	0.074	0.067	0.059	0.055	0.054	0.123	0.101	0.084	0.078	0.069	0.064	0.062	5.17	4.20	3.64	3.30	2.90	2.64	2.59
Insulin	0.065	0.045	0.038	0.039	0.031	0.029	0.054	0.104	0.072	0.066	0.076	0.056	0.050	0.086	3.73	2.41	2.26	2.46	1.86	1.74	2.96
Allowable (%)	41 (91)	41 (91)	41 (91)	42 (93)	43 (96)	44 (98)	44 (98)	40 (89)	40 (89)	40 (89)	40 (89)	41 (91)	43 (96)	43 (96)	38 (84)	40 (89)	40 (89)	40 (89)	41 (91)	42 (93)	44 (98)
Optimal (%)	34 (76)	37 (82)	37 (82)	37 (82)	38 (84)	38 (84)	40 (89)	30 (67)	35 (78)	35 (78)	36 (80)	37 (82)	39 (87)	40 (89)	31 (69)	35 (78)	36 (80)	37 (82)	38 (84)	38 (84)	38 (84)

Table 4 The sample size of the panel and precision of converting RIs [SDR_{LL}, SDR_{UL}, and $CV(b)$].
*1 SDR_{LL} or SDR_{UL} values between 0.125 and 0.25, and above 0.25 are marked by graded gray background. *2 $CV(b)$ values between 5.5% and 11%, and above 11% are marked by graded gray background colors.

Incidentally, the cutoff value of 10% is very close to the allowable limit set in this study for $CV(b)$ computed by the mathematical formula.

However, there was some criticism about judging the allowable level of conversion simply on the basis of r and n. Therefore, in this study, we also evaluated the SE of the converted reference limits LL and UL, $SE(LL)$ and $SE(UL)$, respectively, which were computed based on the mathematical formula we have derived in this study (see Appendix B for its proof). To standardize them, we opted to use their ratio to SD_{RI}, which is approximately 1/4 of the width of the converted RI and corresponds to gross between-individual SD, $\sqrt{SD_I^2+SD_G^2}$. The cutoff values for the ratios SDR_{LL} and SDR_{UL} were set in analogy to setting limits for analytical bias, with SDRs=SE/SD_{RI}≤0.25 as an allowable level and SDRs=SE/SD_{RI} ≤0.125 as an optimal level [9].

At the threshold of allowable precision (SDR_{UL}=0.25), assuming a standardized Gaussian distribution of test values (z) and the converted RIs of LL=–1.96 and of UL=1.96, the 90% CI of UL is computed as z=1.55–2.37 (1.96±1.645×0.25). This implies that when UL is computed as z=1.55 as an extreme case, we would expect a 3.56% absolute increase (or 2.42-fold relative increase) in false-positive cases using the RI: $P(z≥1.55)–P(z≥1.96)$=0.0606–0.0250=0.0356 (or 0.0606/0.025=2.42). Similarly, the UL of z=2.37 as an opposite case gives a 1.61% (0.0250–0.0089) absolute increase (2.81-fold relative increase) of false-negative cases. At the threshold of optimal precision (SDR_{UL}=0.125), the 90% CI of the UL is z=1.75–2.16, and thus, we would expect a 1.51% (0.0401–0.0250) absolute increase (1.60-fold relative increase) in false-positive cases at z=1.75, and 0.96% (0.025–0.0154) absolute increase (1.62-fold relative increase) in false-negative cases at z=2.16.

From this outcome analysis of conversion error, the use of conversions at the optimal level of precision seems better in a practical sense, giving smaller rates of misclassification in using the RI than the larger error rates associated with use of conversions at the allowable level of precision.

An important issue we found in applying SDRs in the judgment of convertibility was that SDRs depended on the distribution pattern of test results for the panel. When the test results showed a symmetrical or Gaussian distribution (|Sk|<0.4), SDR_{LL} and SDR_{UL} were equivalent, but when the distribution was skewed toward the higher values, as is commonly seen in test results for TG, ALT, or GGT, SDR_{UL} tended to become larger relative to SDR_{LL}. Therefore, judgment on convertibility depended solely on the value of SDR_{UL}.

To confirm the appropriateness of the mathematical formulae to derive $CV(b)$ and SDRs, the theoretical values were compared with values derived empirically using the bootstrap method. As expected, both values matched very well when test values used for the method comparison distribute in a symmetrical or Gaussian manner. However, when the distribution was skewed with a long tail toward higher values, the calculated $CV(b)$ and SDRs tended to dissociate from those empirically derived by the bootstrap method.

From the study of log transformation of test results with skewed distributions, it was found that SDR_{UL} computed by the theoretical formula tended to underestimate the value compared with that derived by the bootstrap method. Therefore, for a skewed distribution, it is recommended to calculate the SDR_{UL} after the Gaussian transformation of test results or to estimate it based on the bootstrap method. Meanwhile, the SDR_{LL} for a skewed distribution was found to be overestimated by either of the methods, and thus, it is necessary to convert it into a Gaussian distribution for more reliable derivation of SDR_{LL}.

In contrast, the calculated $CV(b)$ appears to be more robust and less affected by skewness than the SDRs. Therefore, $CV(b)$ appears to be an index of choice in judging the practicability of converting the RI based on the regression line.

No rational theory for setting the threshold level of $CV(b)$ is available at present. However, a very close relationship was found in this study between $CV(b)$ and SDRs. We empirically derived a linear equation between $CV(b)$ and aveSDR (average of SDR_{LL} and SDR_{UL}) from the entire set of data on method comparison for all 45 analytes. The equation can be approximated as SDR=0.22 $CV(b)$, and SDR=0.25 and SDR=0.125 are equivalent to $CV(b)\approx11\%$ and $CV(b)\approx5.5\%$ in judging the allowable and optimal precision, respectively. It is important to note that this relationship holds true only when the values of the mean±1.96SD, as LL_0 and UL_0, are converted to obtain LL and UL (see Appendix B).

It is obvious that the more specimens there are in the panel, the higher the precision of the conversion. From the results of our simulation study shown in Table 4, we found that a majority of the analytes had optimum precision for conversion either judged by $CV(b)$ or by SDRs with $n\geq60$. The precision declines gradually over the range of $30\leq n\leq50$. However, the average $CV(b)$ and SDRs increase more sharply with $n\leq30$.

The analytes we plan to measure in the worldwide study are numerous, and they are heterogeneous in their behavior in the method comparison. Therefore, there is no particular best sample size to cope with all situations. Linnet evaluated the sample size necessary for method comparison studies by simulation [12] and reported that the range (LL–UL) of values under comparison is very critical in determining the sample size. He concluded that a range ratio (maximum value divided by minimum value) of <2.0 requires a larger sample size for proper comparison. In our study, among the analytes whose $CV(b)$ or SDR was relatively high, TP, Alb, Na, Cl, and Ca showed $r<0.9$, and the ratio of UL to LL, which corresponds to the range ratio, was <1.3. Therefore, the test results for those analytes must be compared using a large number of specimens, which may be an impractical number (n) as predicted from the theoretical formula for $CV(b)$; i.e., assuming r=0.5, 0.6, 0.7, or 0.8, to achieve the allowable limit of $CV(b)$=11%, the n required is 250, 149, 99, or 49; for the optimal limit of $CV(b)$=5.5%, the n required is 994, 590, 346, or 188. As a result, these analytes are likely to be excluded from the comparison when the observed r is <0.8. We should rather standardize the test results of each testing center and compare the results without making an attempt to adjust values by method comparison.

The sample size of 40 that we adopted is equivalent to that stipulated in the EP9-A2 for *Method Comparison and Bias Estimation Using Patient Samples* [13]. The measurement of each specimen in duplicate is also suggested, in consideration of within-day variations of the assay. However, we consider it more important to measure the panel on separate days after dividing them into multiple parts to include both within- and between-day variations in the test results. In fact, we observed occasional bias in the repeated measurement of the panel 1 week apart, and the between-day components of error outweighed the within-day components. Therefore, we strongly recommend that the 40 sera be measured on 4–8 separate days in the worldwide study.

At the end of this article, we have to admit that our strategy to use the panel of sera for the alignment of reference values across the countries remains largely exploratory, although we provided theoretical formulation of the statistical approach for converting the values. Therefore, the approach may require a reevaluation from the practical point of view after their implementation in the ongoing worldwide study.

Acknowledgments: The research funds used included a Scientific Research Fund (No. 21406015: 2009–2011)

provided by the Japan Society for the Promotion of Science (JSPS), which was used for the preparation of the panel of sera, purchase of CRMs, and logistics; the reagents required for testing all the target analytes in Uludag University were provided by Abbott Co. and those for testing in Japan were provided by Beckman Coulter Japan; the reagents used by the Mayo Clinic and ARUP Laboratories were procured with institutional research funds; some of the immunoturbidometric reagents for the study in Japan were provided by Nittobo Co. We are very grateful to the members of the Scientific Division of the IFCC, especially Prof. Philippe Gillery, for critical review and editing of this manuscript.

Conflict of interest statement

Authors' conflict of interest disclosure: The authors stated that there are no conflicts of interest regarding the publication of this article. Research funding played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.

Received October 26, 2012; accepted March 23, 2013

Appendix A

Derivation of standard error of the slope *b* estimated by the major axis regression

Derivation of standard error of slope b [SE(b)] is shown below on the basis of additivity of the normal distribution.

The notations used in the derivation are as follows with regard to n sets of paired observations x_i, y_i(i=1, 2, ⋯, n).

n: sample size (i=1, 2, ⋯, n)

$$S_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$S_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$S_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

r: Pearson's correlation coefficient, and $r=S_{xy}/\sqrt{S_{xx}S_{yy}}$

s_{LS}: SD around the least-square regression line (along y-axis) and $s_{LS}=\sqrt{(1-r^2)S_{yy}/(n-2)}$.

The linear regression equation based on the major axis regression is expressed as $y=\bar{y}+b(x_i-\bar{x})$ where b represent the slope of the major axis regression and is expressed as $\sqrt{S_{xx}/S_{yy}}$, while the slope of the least square regression is denoted as b_{LS} which is expressed as S_{xy}/S_{xx}.

Assuming that a variable x follows normal distribution, the additivity of the normal distribution allows the following operations, in which $\bar{x}$ and s^2 denote the mean and variance of x, and a, a_1 and a_2 indicating constant values.

variance of value x: $Var(x)=s^2$

vairance of value ax: $Var(ax)=a^2Var(x)=a^2s^2$

variance of value a_1x+a_2x: $Var(a_1x+a_2x)=Var(a_1x)+Var(a_2x)=\left(a_1^2+a_2^2\right)s^2$.

The variance of b_{LS} [Var(b_{LS})] is well known (citation 5) and the variance of b [Var(b)] can be derived using the information as follows:

$$Var(b_{LS})=\frac{s_{LS}^2}{S_{xx}}=\left(\frac{(1-r^2)S_{yy}}{n-2}\right)/S_{xx}$$

$$\begin{aligned}Var(b) &=Var\left(\sqrt{\frac{S_{yy}}{S_{xx}}}\right)=Var\left(\frac{\sqrt{S_{xx}\cdot S_{yy}}}{S_{xx}}\right) \quad \Leftarrow r=\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\\ &=Var\left(\frac{1}{S_{xx}}\times\frac{S_{xy}}{r}\right)=\frac{1}{r^2}Var\left(\frac{S_{xy}}{S_{xx}}\right) \quad \Leftarrow \frac{S_{xy}}{S_{xx}}=b_{LS}\\ &=\frac{1}{r^2}Var(b_{LS})=\frac{1}{r^2}\left(\frac{1-r^2}{n-2}\right)\frac{S_{yy}}{S_{xx}} \quad \Leftarrow Var(b_{LS})\\ &=\frac{(1-r^2)S_{yy}}{(n-2)S_{xx}},\ \frac{S_{xx}}{S_{yy}}=b^2=\frac{b^2}{r^2}\left(\frac{1-r^2}{n-2}\right)\end{aligned}$$

Therefore,

$$SE(b)=\sqrt{Var(b)}=\frac{b}{r}\sqrt{\frac{1-r^2}{n-2}}$$

$$CV(b)=\frac{SE(b)}{b}\times 100=\frac{100}{r}\sqrt{\frac{1-r^2}{n-2}}$$

Appendix B

Derivation of standard error of converted reference limits (LL, UL) by use of the major axis regression line.

In general, the variance of y [Var(y)], predicted from x by the major axis regression formula $y=\bar{y}+b(x-\bar{x})$ can be derived as shown below using the following notations: $\bar{x}$ and s_x denoting mean and SD of observed values x; s_{MA} denoting SD around the major axis regression line; s_{LS} denoting SD around the least-square regression line:

$$Var(y)=Var(\bar{y}+b(x-\bar{x}))$$
$$=Var(\bar{y})+(x-\bar{x})^2Var(b) \quad \Leftarrow Var(b)=\frac{Var(b_{LS})}{r^2}$$
$$=\frac{s_{MA}^2}{n}+(x-\bar{x})^2\frac{Var(b_{LS})}{r^2} \quad \Leftarrow s_{MA}^2=\frac{2}{1+r}s_{LS}^2 \text{ [a]}$$
$$=\frac{2}{(1+r)n}s_{LS}^2+(x-\bar{x})^2\frac{s_{LS}^2}{r^2S_{xx}} \quad \Leftarrow S_{xx}=(n-1)s_x^2$$
$$=\left(\frac{2}{(1+r)n}+\frac{(x-\bar{x})^2}{r^2(n-1)s_x^2}\right)s_{LS}^2$$

Assuming that UL_0 was obtained as $\bar{x}+1.96s_x$, we set $x=\bar{x}+1.96s_x$ in the above equation and derive $SE(UL)$ as folows [$SE(LL)$ can be derived in the same manner and therefore not shown here]:

$$Var(UL)=\left(\frac{2}{(1+r)n}+\frac{(\bar{x}+1.96s_x-\bar{x})^2}{r^2S_{xx}}\right)s_{LS}^2$$
$$=\left(\frac{2}{(1+r)n}+\frac{3.84s_x^2}{r^2(n-1)s_x^2}\right)s_{LS}^2$$
$$=\left(\frac{2}{(1+r)n}+\frac{3.84}{r^2(n-1)}\right)s_{LS}^2$$

$$SE(UL)=\sqrt{Var(UL)}=s_{LS}\sqrt{\frac{2}{(1+r)n}+\frac{3.84}{r^2(n-1)}}$$

$$s_{MA}=\sqrt{\frac{\Sigma(y_i-[\bar{y}+b(x_i-\bar{x})])^2}{n-2}} \quad \Leftarrow \text{from definition of } s_{MA}$$
$$=\sqrt{\frac{\Sigma(y_i-\bar{y})^2-2b(x_i-\bar{x})(y_i-\bar{y})+b^2(x_i-\bar{x})^2}{n-2}}$$
$$=\sqrt{\frac{S_{yy}-2bS_{xy}+b^2S_{xx}}{n-2}} \quad \Leftarrow b^2=S_{yy}/S_{xx}$$
$$=\sqrt{\frac{S_{yy}-2bS_{xy}+S_{yy}}{n-2}}$$
$$=\sqrt{\frac{2(S_{yy}-bS_{xy})}{n-2}}$$

$$\frac{s_{MA}}{s_{LS}}=\sqrt{\frac{2(S_{yy}-bS_{xy})}{S_{yy}-b_{LS}S_{xy}}} \quad \Leftarrow b=\sqrt{\frac{S_{yy}}{S_{xx}}},\ b_{LS}=\sqrt{\frac{S_{xy}}{S_{xx}}}$$
$$=\sqrt{\frac{2(S_{yy}-\sqrt{S_{yy}/S_{xx}}S_{xy})}{S_{yy}-S_{xy}^2/S_{xx}}}$$
$$=\sqrt{\frac{2(S_{xx}S_{yy}-\sqrt{S_{yy}S_{xx}}S_{xy})}{S_{xx}S_{yy}-S_{xy}^2}} \quad \Leftarrow r=\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$
$$=\sqrt{\frac{2(S_{xy}^2/r^2-S_{xy}^2/r)}{S_{xy}^2/r^2-S_{xy}^2}}$$
$$=\sqrt{\frac{2(1-r)}{1-r^2}}=\sqrt{\frac{2}{1+r}}$$

Therefore,

$$s_{MA}=\sqrt{\frac{2}{1+r}}\cdot s_{LS}$$

[a]The proof of $s_{MA}^2=2/(1+r)\cdot s_{LS}^2$

$$s_{LS}=\sqrt{\frac{\Sigma(y_i-[\bar{y}+b_{LS}(x_i-\bar{x})])^2}{n-2}} \quad \Leftarrow \text{from definition of } s_{LS}$$
$$=\sqrt{\frac{\Sigma(y_i-\bar{y})^2-2b_{LS}(x_i-\bar{x})(y_i-\bar{y})+b_{LS}^2(x_i-\bar{x})^2}{n-2}}$$
$$=\sqrt{\frac{S_{yy}-2b_{LS}S_{xy}+b_{LS}^2S_{xx}}{n-2}} \quad \Leftarrow b_{LS}=S_{xy}/S_{xx}$$
$$=\sqrt{\frac{S_{yy}-b_{LS}S_{xy}}{n-2}}$$

References

1. Ozarda Y, Ichihara K, Barth JH, Klee G. The protocol and standard operation procedures for common use in the worldwide multicenter study on reference values. Clin Chem Lab Med 2013;51:1027–40.
2. Ludbrook J. Comparing methods of measurement. Clin Exp Pharmacol Physiol 1997;24:193–203.
3. McArdle BH. The structural relationship: regression in biology. Can J Zool 1988;66:2329–39.
4. Bohonak AJ. RMA: software for reduced major axis regression. Available from: http://www.bio.sdsu.edu/pub/andy/RMA.html. Accessed 23 April, 2012.
5. Sokal RR, Rohlf FJ. Linear regression (type model II regression). In: Biometry: the principles and practice of statistics in biological research, 3rd ed. New York, USA: WH Freeman, 1995:541–9.
6. Draper NR, Smith H. Applied regression analysis, 3rd ed. New York: John Wiley & Sons, 1998.
7. Linnet K. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. Clin Chem 1998;44:1024–31.
8. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge: Cambridge University Press, 1997.
9. Fraser CG. Biological variation: from principle to practice. Washington, DC: AACC Press, 2001.
10. Ichihara K, Ceriotti F, Tam TH, Sueyoshi S, Poon PM, Thong ML, et al. The Asian project for collaborative derivation of reference intervals: (1) strategy and major results of standardized analytes. Clin Chem Lab Med 2013. DOI: 10.1515/cclm-2012-0421 [Epub ahead of print].
11. Ichihara K, Ceriotti F, Kazuo M, Huang YY, Shimizu Y, Suzuki H, et al. The Asian project for collaborative derivation of reference intervals: (2) results of non-standardized analytes and transference of reference intervals to the participating laboratories on the basis of cross-comparison of test results. Clin Chem Lab Med 2013. DOI: 10.1515/cclm-2012-0422 [Epub ahead of print].
12. Linnet K. Necessary sample size for method comparison studies based on regression analysis. Clin Chem 1999;45: 882–94.
13. NCCLS. Method comparison and bias estimation using patient samples: approved guideline, 2nd ed. NCCLS document EP9-A2. Wayne, PA: Clinical and Laboratory Standards Institute, 2002.