PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2003; **2**: 253–261 (DOI:10.1002/pst.052)

An improved concordance correlation coefficient

MAIN PAPER

Jason J. Z. Liao*,†
Merck Research Laboratories, Merck & Co. Inc., West Point, PA 19486, USA

It is often required to compare two measurements in medicine and other experimental sciences. This problem covers a broad range of data, and examples can be found in different industries. In this paper, a new index on measuring agreement is proposed, which is similar to Lin's concordance correlation coefficient but derived from a criterion which is more conceptually appealing and which offers improvements. An example is used to demonstrate the benefit of using the new index. Copyright © 2003 John Wiley & Sons Ltd.

1. INTRODUCTION

It is often a requirement in medicine and other experimental sciences to assess the agreement between two measurements made by devices, laboratories or raters, and the agreement between two clinical treatments or two methods. The agreement problem covers a broad range of data, and examples can be found in different industries. In addition to many examples from medical-related industries, interesting applications can be found in Vonesh and Chinchilli [1], MacLennan [2], Guerrero and Roshwalb [3], Hofstee and Zegers [4], Krippendorff [5], and Tanner and Young [6] among others.

The agreement problem has a long history and can be traced back to the first approach in 1889 and 1901 by Pearson, who proposed the correlation coefficient to measure the agreement. Later, other methods, such as the paired t-test, regression analysis and intraclass correlation coefficient (ICC), were developed. However, some of these methods may lead to misleading conclusions. Pearson's correlation coefficient ρ, and therefore the coefficient of determination ρ^2, measures the linear relationship by measuring how far observations deviate from the best-fit line. Consequently, $\rho(\rho)^2$ does not distinguish between situations where two measurements strongly agree (high correlation about the line of identity $y=x$) and those where a strong linear relationship exists but the measurements do not agree. The paired t-test compares the mean value of two measurements, which can fail to indicate agreement at the individual measurement level. Regression analysis tests the departure from the intercept equal to 0 and the slope equal to 1, which can reject a reasonably good agreement if the residual errors are small. The coefficient of variation (CV) and intraclass correlation coefficient interchange duplicate readings where duplicate readings are actually two distinct measurements [7], and therefore they can be misleading. In addition, the CV and ICC assume a fixed mean, which is usually not

*Correspondence to: Jason J. Z. Liao, Merck Research Laboratories, Merck & Co. Inc., West Point, PA 19486, USA.
†E-mail: jason_liao@merck.com

the case in practice. Bland and Altman [8] proposed a graphic tool that plots the difference against the mean of the two paired measurements. For the case where a mixture of fixed, proportional bias and proportional error occurs, the Bland and Altman's graphic tool is not appropriate [9]. In addition, the mean of the two measurements used in their approach is always a random variable even if one of the two measurements comes from a 'gold' standard.

As mentioned in Lin and Torbeck [10], all of these approaches give the right answer to a wrong question. The right question one should address is 'are the two measurements precise and accurate across a range and tightly clustered along the 45° line through the origin?' That is, one should divide the agreement assessment into two steps. First, a precision step assesses how close the observations are to the regression line. Second, an accuracy step assesses how close the regression line and the target line (the line of identity) are to each other. In an attempt to give an answer to the right question, Lin [7] proposed a concordance correlation coefficient (denoted as $CCC_{\text{Lin}} = \rho C_a^{\text{Lin}}$) under the assumption that the observations are from a bivariate normal distribution. This index combines the measure of precision (ρ) and measure of accuracy ($C_a^{\text{Lin}} = 2\sigma_1\sigma_2/[\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2]$) together. However, since C_a^{Lin} does not include ρ as a component, the accuracy is the same no matter how the two measurements are correlated, as long as their corresponding mean and variance are the same. Other concerns such as unexplainable or misleading results from CCC_{Lin} and C_a^{Lin}can be found in Muller and Buttner [11], Nickerson [12], and Liao and Lewis [13]. Some of these problems are shared with other existing correlation approaches.

Because of this controversy, a new agreement index is proposed in Section 2, which gives an improved assessment of agreement. Its properties are studied in Section 3. In Section 4, the new index is compared with CCC_{Lin} by their theoretical values and by means of a simulation study. An example is presented in Section 5 to demonstrate the benefit of using the proposed index. A summary and a discussion follow in Section 6.

2. THE NEW CONCORDANCE CORRELATION COEFFICIENT

The criterion used by Lin [7] to derive $CCC_{\text{Lin}} = 2\rho\sigma_1\sigma_2/[\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2]$ is the squared perpendicular distance $(Y_1 - Y_2)^2$ of any paired observation (Y_1, Y_2) to the line of identity, which represents a perfect agreement. Unfortunately, the squared perpendicular distance only measures how close the regression line is to the line of identity, i.e. the accuracy. When this accuracy part is fixed, the regression line can still change direction, as shown in Figure 1(a). In other words, the precision, which measures how close the observations are to the regression line, is not fixed. Since only one point (i.e., one paired observation (Y_1, Y_2)) cannot determine a regression line, the observations can be around either the solid line or the dotted line. Therefore, another point needs to be introduced to determine the regression line.

Let us use the same assumption as Lin [7] that pairs of samples (Y_{1i}, Y_{2i}), $i = 1, \ldots, n$, are independently selected from a bivariate normal population with mean μ_1 and μ_2, and covariance matrix

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Instead of Lin's [7] expected distance $E(Y_2 - Y_1)^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 + (\mu_2 - \mu_1)^2$ for any pair (Y_1, Y_2) to the line of identity, the expected square value for an area is used. The area is formed by

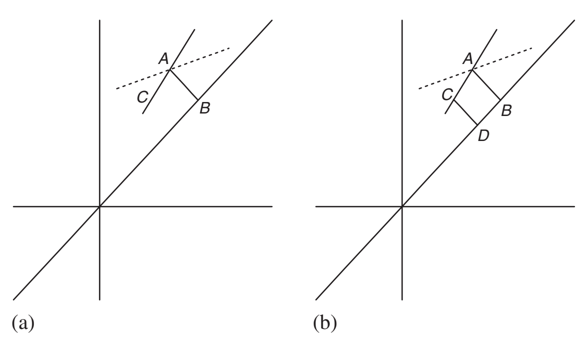

Figure 1. Comparison of the two criteria: (a) Lin's criteria; (b) new criterion.

any *two* paired observations (Y_{11}, Y_{21}) and (Y_{12}, Y_{22}) to the line of identity:

$$\left[\frac{(Y_{11}-Y_{12})^2}{\sigma_1^2}+\frac{(Y_{21}-Y_{22})^2}{\sigma_2^2}\right](Y_{11}-Y_{21})^2$$

The area $ABDC$ in Figure 1(b) leads to the solid line in Figure 1(a) which no longer has the ambiguity of Lin's criterion on which line to use. Thus, two points (i.e., two paired observations) are used to determine both the regression line and the distance to the line of identity. The expectation of the area is

$$E\left\{\left[\frac{(Y_{11}-Y_{12})^2}{\sigma_1^2}+\frac{(Y_{21}-Y_{22})^2}{\sigma_2^2}\right](Y_{11}-Y_{21})^2\right\} = (3+\rho^2)(\sigma_1^2+\sigma_2^2) - 8\rho\sigma_1\sigma_2 + 2\Delta^2, \quad (1)$$

where $\Delta = \mu_2 - \mu_1$.

Remarks

1. The area $ABDC$ combines the two steps mentioned at the beginning of this section. $(Y_{11}-Y_{21})^2$ (AB in Figure 1(b)), which is proportional to the squared perpendicular deviation from the line of identity and was used by Lin [7], measures the accuracy. $(Y_{11}-Y_{12})^2/\sigma_1^2 + (Y_{21}-Y_{22})^2/\sigma_2^2$ (AC in Figure 1(b)) is the standardized distance of two pairs of observations and measures the precision. The latter was overlooked by Lin [7]. Given this area, the relationship between the two measurements can be determined. The area $ABDC$ is a more conceptually appealing criterion than the distance used in Lin [7].
2. To derive the expectation of the area, the following properties were used.

   (a) Smoothness property, $E(Y) = E[E(Y|X)]$.
   (b) If

   $$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right),$$

   then

   $$Y|X \sim N(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(X-\mu_1), \sigma_2^2(1-\rho^2));$$

   (c) If $X \sim N(0, s^2)$, then $E(X^4) = 3s^4$.
3. If Y_1 and Y_2 have a perfect match, then $\mu_1 = \mu_2, \sigma_1 = \sigma_2, \rho = 1$; hence the expectation in equation (1) is 0.

In order to create an index on the scale from -1 to 1, consider an expression of the agreement index,

$$r_\rho = \frac{\text{Expected decreased area for } \rho \text{ units}}{\begin{matrix}\text{(Expected area} \\ \rho=+1)\end{matrix} + \begin{matrix}\text{(Expected decreased} \\ \text{area for 1 unit)}\end{matrix}}$$
$$= \rho\frac{4\sigma_1\sigma_2 - \rho(\sigma_1^2+\sigma_2^2)}{(2-\rho)(\sigma_1^2+\sigma_2^2)+\Delta^2}$$
$$= \rho\frac{4\tau - \rho(1+\tau^2)}{(2-\rho)(1+\tau^2)+\delta^2},$$

where $\tau = \sigma_2/\sigma_1$ and $\delta = \Delta/\sigma_1$. Let

$$A_\rho = \frac{4\sigma_1\sigma_2 - \rho(\sigma_1^2+\sigma_2^2)}{(2-\rho)(\sigma_1^2+\sigma_2^2)+\Delta^2} = \frac{4\tau - \rho(1+\tau^2)}{(2-\rho)(1+\tau^2)+\delta^2}$$

Then A_ρ is a measure of the accuracy. This measure depends not only on the scale ratio and location shift relative to the 'gold' method (typically with smaller measurement error), but also on the correlation coefficient ρ.

Remarks

1. The expected area in equation (1) is a quadratic function of ρ but the expected distance in Lin [7] is a linear function of ρ. If the above approach of mapping the expected area to $[-1,1]$ is applied to Lin's [7] expected distance, one could obtain CCC_{Lin}.
2. Notice that the accuracy statistic C_a^{Lin} in Lin [7] does not include ρ as a component, but the accuracy index $A_\rho \in [0,1]$, defined above, does. Conceptually, the accuracy measurement should be a function of ρ because of the appearance of ρ in the conditional

distribution of $Y_2|Y_1$. Other arguments and examples can be found in Liao and Lewis [13]. Therefore, the statistic developed here is more reasonable.

3. If the two measurements have the same mean and variance, i.e., $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, then both measures of accuracy are equal to 1, and the measures of agreement are equal to ρ.

3. INFERENCE

A_r and γ_ρ can be estimated by replacing the parameters with their corresponding sample values as follows:

$$\hat{A}_\rho = \frac{4S_1S_2 - \hat{\rho}(S_1^2 + S_2^2)}{(2-\hat{\rho})(S_1^2 + S_2^2) + (\bar{Y}_1 - \bar{Y}_2)^2}$$

and

$$\hat{r}_\rho = \hat{\rho}\hat{A}_\rho$$

where

$$\hat{\rho} = \frac{S_{12}}{S_1S_2}, \bar{Y}_1 = \frac{1}{n}\sum_{i=1}^{n} Y_{1i}, \bar{Y}_2 = \frac{1}{n}\sum_{i=1}^{n} Y_{2i},$$

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2, S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_{2i} - \bar{Y}_2)^2,$$

$$S_{12} = \frac{1}{n}\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)$$

The proof of the following lemma is straightforward.

Lemma 1 *Let* $v^{\mathrm{T}} = (v_1, v_2, v_3, v_4) = (S_1^2, S_2^2, S_{12}, \bar{Y}_2 - Y_1)$. *Then* v *has an asymptotic normal distribution with mean* $E(v) = (\sigma_1^2, \sigma_2^2, \rho\sigma_1\sigma_2, \mu_2 - \mu_1)$ *and variance* $n^{-1}\Sigma$, *where*

$$\Sigma = \begin{pmatrix} 2\sigma_1^4 & 2\rho^2\sigma_1^2\sigma_2^2 & 2\rho\sigma_1^3\sigma_2 & 0 \\ 2\rho^2\sigma_1^2\sigma_2^2 & 2\sigma_2^4 & 2\rho\sigma_1\sigma_2^3 & 0 \\ 2\rho\sigma_1^3\sigma_2 & 2\rho\sigma_1\sigma_2^3 & (1+\rho^2)\sigma_1^2\sigma_2^2 & 0 \\ 0 & 0 & 0 & \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 \end{pmatrix}$$

Now,

$$\hat{\gamma}_\rho = \frac{4v_3v_1v_2 - v_3^2(v_1 + v_2)}{[2v_1v_2 - v_3\sqrt{v_4v_2}](v_1 + v_2) + v_4^2v_1v_2}$$

From Lemma 1, $g(v)$ is asymptotic normally distributed with mean $g(E(v))$ and variance $n^{-1}d\Sigma d'$, where

$$d = \left(\left.\frac{\partial g}{\partial v_1}\right|_{v=E(v)}, \ldots, \left.\frac{\partial g}{\partial v_4}\right|_{v=E(v)} \right)$$

Then the following results are obtained.

Theorem 1 Let

$$Z(\gamma_\rho) = \tanh^{-1}(\gamma_\rho) = \frac{1}{2}1n\frac{1+\gamma_\rho}{1-\gamma_\rho}$$

Then

(a) *$(Z(\hat{\gamma}_\rho)$ has an asymptotic normal distribution with mean $Z(\gamma_p)$ and variance* $\frac{1}{n}\frac{1}{(1-\gamma_\rho^2)^2}V_z$

(b) *$\hat{\gamma}_\rho$ has an asymptotic normal distribution with mean γ_ρ and variance V_z/n, where $V_z = c^{\mathrm{T}}\Sigma c, c = (c_1, c_2, c_3, c_4)^{\mathrm{T}}, \Sigma$ is from Lemma 1 and*

$$c_1 = \frac{4\rho\sigma_1\sigma_2^3 - \rho^2\sigma_1^2\sigma_2^2}{\Omega} - \frac{\gamma_\rho^2}{\sigma_1^2} - \gamma_\rho\frac{2\sigma_1^2\sigma_2^2 + 0.5\rho\sigma_1^2\sigma_2^2 + 1.5\rho\sigma_2^4}{\Omega},$$

$$c_2 = \frac{4\rho\sigma_1^3\sigma_2 - \rho^2\sigma_1^2\sigma_2^2}{\Omega} - \frac{\gamma_\rho^2}{\sigma_2^2} - \gamma_\rho\frac{2\sigma_1^2\sigma_2^2 + 0.5\rho\sigma_1^2\sigma_2^2 + 1.5\rho\sigma_1^4}{\Omega},$$

$$c_3 = \frac{4\sigma_1^2\sigma_2^2 - 2\rho\sigma_1\sigma_2^3 - 2\rho\sigma_1^3\sigma_2}{\Omega} + \gamma_\rho\frac{\sigma_1\sigma_2(\sigma_1^2 + \sigma_2^2)}{\Omega},$$

$$c_4 = -\gamma_\rho\frac{2\sigma_1\sigma_2(\mu_2 - \mu_1)}{\Omega}$$

and $\Omega = \sigma_1^2\sigma_2^2[(2-\rho)(\sigma_1^2 + \sigma_2^2) + (\mu_2 - \mu_1)^2]$.

4. COMPARISON

When $\rho = 1$ and $\sigma_1 = \sigma_2$ or $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, then $\gamma_\rho = CCC_{\mathrm{Lin}}$. However, there is no general relationship between γ_ρ and CCC_{Lin} or between A_ρ and C_a^{Lin}. Note that it is always true that $CCC_{\mathrm{Lin}} \leqslant ICC$.

In order to evaluate how CCC_{Lin} and γ_ρ perform, the difference $CCC_{\mathrm{Lin}} - \gamma_\rho$ is calculated for $\mu_1 = \mu_2$ and most of the practical situations where σ_2/σ_1 and ρ are in [0.5,1]. This difference is plotted in Figure 2, which indicates that CCC_{Lin} and γ_ρ are very similar when σ_2/σ_1 is around 1. At a fixed ρ, the difference $CCC_{\mathrm{Lin}} - \gamma_\rho$ increases as σ_2/σ_1 decreases. However, the increase rate is different at different ρ. If σ_2/σ_1 is fixed, then the difference $CCC_{\mathrm{Lin}} - \gamma_\rho$ is an increasing function of ρ. The biggest difference occurs at $\sigma_2/\sigma_1 = 0.5$ and $\rho = 1$, where $CCC_{\mathrm{Lin}} = 0.8$ while $\gamma_\rho = 0.6$. This indicates that γ_ρ value is more reasonable and CCC_{Lin} is too big since $\sigma_1 = 2 \times \sigma_2$ and $Y_2|Y_1 = 0.5Y_1 + 0.5\mu_1$, which deviates a lot from the ideal agreement case $Y_2|Y_1 = Y_1$.

Lin and Torbeck [10] proposed a threshold value of $CCC_{\mathrm{Lin}} \geqslant 0.97$ as indicating good agreement, which is too high to be reasonable. For the corresponding agreement index of the categorical data, there is a useful benchmark cut-off for the Kappa statistic: 0.21–0.40 being a fair agreement, 0.41–0.60 being a moderate agreement, 0.61–0.80 being a substantial agreement and 0.81–1.00 being an almost perfect agreement [14]. Figure 2 indicates that CCC_{Lin} usually gives a higher value for most practical situations. This is a possible reason why Lin and Torbeck [10] proposed a high threshold CCC_{Lin} value as a good agreement.

In order to obtain a more clearer picture regarding these two indices, simulations are conducted for observations from the following two cases:

1. $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 2.5 \\ 3.5 \end{pmatrix}, \begin{pmatrix} 1^2 & 0.9 \times 1 \times 2 \\ 0.9 \times 1 \times 2 & 2^2 \end{pmatrix}\right)$;

2. $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 3.5 \\ 3.5 \end{pmatrix}, \begin{pmatrix} 1^2 & 0.9 \times 1 \times 2 \\ 0.9 \times 1 \times 2 & 2^2 \end{pmatrix}\right)$.

Case 1 has different means and different variances for the two variables, while case 2 has different variances. The summary statistics are reported in Table I for 1000 samples. Table I shows that CCC_{Lin} tends to give a higher value than γ_ρ. The improvement of the agreement of case 2 is due to the equality of the mean values of the two variables. The histograms of $CCC_{\mathrm{Lin}} - \gamma_\rho$ are reported in Figure 3, which shows a very symmetric distribution.

5. ASSAY TRANSFER DATA

In the pharmaceutical industry, an assay is often transferred from one laboratory to another, as when a well-developed assay is transferred from a research laboratory to a manufacturing laboratory. Before a new laboratory can start to perform a new assay, it must be validated by assessing whether the new laboratory obtains results similar to the current laboratory. Hence, the agreement of two laboratories must be evaluated.

Consider an assay transfer study where the assay was validated in the current laboratory and there is no analyst effect. To transfer the assay to a new laboratory, three qualified analysts were randomly chosen from each laboratory and randomly paired to conduct an experiment for

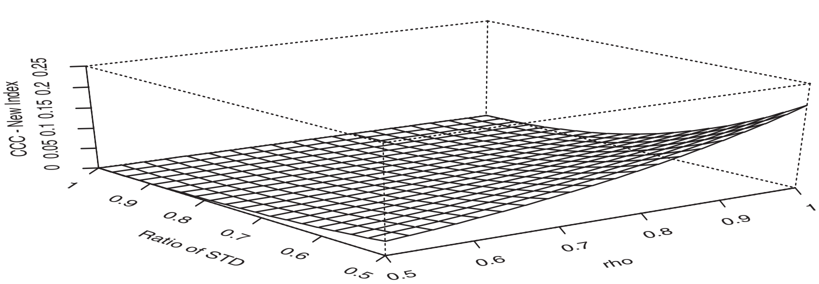

Figure 2. Comparison of CCC_{Lin} and γ_ρ for $\mu_1 = \mu_2$.

Pharmaceutical STATISTICS

Table I. Simulation comparison of γ_ρ with CCC_{Lin}.

	Case 1			Case 2		
	CCC_{Lin}	γ_ρ	$CCC_{\mathrm{Lin}}-\gamma_\rho$	CCC_{Lin}	γ_ρ	$CCC_{\mathrm{Lin}}-\gamma_\rho$
Minimum	0.453	0.366	0.054	0.619	0.450	0.08
Maximum	0.687	0.599	0.180	0.790	0.674	0.203
Mean	0.596	0.481	0.155	0.717	0.570	0.147
Median	0.598	0.481	0.155	0.718	0.571	0.146
2.5% Quartile	0.522	0.413	0.081	0.633	0.500	0.113
97.5% Quartile	0.622	0.547	0.149	0.765	0.644	0.182
Standard Deviation	0.037	0.035	0.018	0.026	0.037	0.019
True Regression Line	$Y_2 \mid Y_1 = -1 + 1.8\,Y_1$			$Y_2 \mid Y_1 = -2.8 + 1.8\,Y_1$		

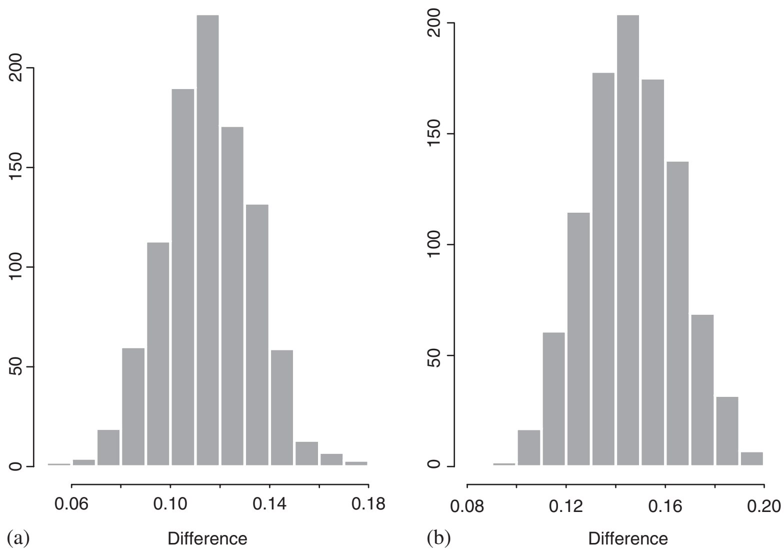

Figure 3. Distribution of $CCC_{\mathrm{Lin}}-\gamma_\rho$: (a) case 1; (b) case 2.

nine different samples, where the samples used in each laboratory were identical. There were a total of 27 results from each laboratory. Figure 4 reports the assay transfer data with the best-fit least-squares regression line and the line of identity. For convenience, Table II lists both the raw data and the summary statistics of the data.

For this assay transfer data set, Pearson's $\hat{\rho} = 0.923$, $C\hat{C}C_{\mathrm{Lin}} = 0.528$, $\hat{C}_a^{\mathrm{Lin}} = 0.572$ and the best-fit line was $Lab_{\mathrm{new}} = 63.94 + 0.29 \times Lab_{\mathrm{old}}$, while $\hat{\gamma}_\rho = 0.19$, and $\hat{A}_\rho = 0.206$. The approximate 95% confidence intervals were [0.450, 0.598] and [0.149, 0.231] for CCC_{Lin} and γ_ρ, respectively. The statistics CCC_{Lin} and C_a^{Lin} gave much larger values than γ_ρ and A_ρ, respectively. Both confidence intervals were based on the asymptotic normality of the Z-transformation in order to improve the normality approximation and the convergence rate. This was then transformed back to the original agreement index. The best-fit line, $Lab_{\mathrm{new}} = 63.94 + 0.29 \times Lab_{\mathrm{old}}$, was far away from the line of identity, $Lab_{\mathrm{new}} = 0 + 1 \times Lab_{\mathrm{old}}$. The accuracy should be very low but $\hat{C}_a^{\mathrm{Lin}} = 0.572$ still gave a high value, while the new index is $\hat{A}_\rho = 0.206$.

Table II indicates that the two laboratories produced the same mean values but did not have

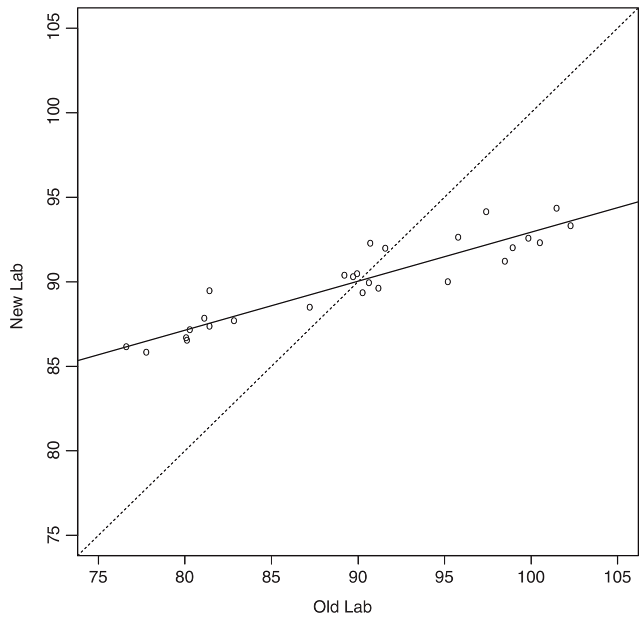

Figure 4. Scatterplot of the assay transfer data (solid line is best-fit line; dotted line is line of identity).

a good agreement, even though they had a very high correlation coefficient. The disagreement appears to result primarily from the less variable performance of the new laboratory. Despite the lack of a high degree of agreement between the two laboratories, people are likely to further evaluate the assay techniques in the new laboratory because of its lower variabiity. If the better performance of the new laboratory is confirmed, the procedure used at the new laboratory will become the preferred assay technique.

There were six possible pairs for the three analysts from each laboratory. The results for remaining five possible pairs were (0.529, 0.189), (0.493, 0.213), (0.491, 0.215), (0.482, 0.220) and (0.479, 0.221) for $(C\hat{C}C_{\text{Lin}}, \hat{r}_\rho)$, respectively. These results give conclusions similar to the first pairing of analysts as detailed above.

6. SUMMARY AND DISCUSSION

In this paper, a new agreement statistic γ_ρ was derived from a more conceptually appealing criterion than the one by Lin [7] to measure the agreement of two measurements. Lin's statistic, CCC_{Lin} and the new statistic, γ_ρ have a similar structure. When the two measurements have similar variations, the new index γ_ρ tends to give a similar value to CCC_{Lin}, but in most practical situations where the two measurements do not have a similar variation, the new index γ_ρ gives better results than CCC_{Lin} does.

Since A_ρ is the measure of accuracy, i.e., it measures how far away the best-fit line deviates, from the line of identity, one can estimate A_ρ in another way. Considering the linear regression of Y_2 on Y_1, one has

$$Y_2|Y_1 = \alpha + \beta Y_1 + \varepsilon$$

where ε has mean 0 and variance σ^2, and ε and Y_1 are independent. One can estimate μ_1, σ_1 and ρ as usual but estimate μ_1 and σ_2 as follows:

$$\hat{\mu}_2 = E(Y_2) = \hat{\alpha} + \hat{\beta}E(Y_1) = \bar{Y}_2$$

$$\hat{\sigma}_2^2 = \hat{\beta}^2\text{var}(Y_1) + \hat{\sigma}^2 = \hat{\beta}^2 S_1^2 + \hat{\sigma}^2$$

Table II. Raw data and summary statistics of the assay transfer data.

		Old Lab	New Lab
Raw Data	1	91.300	89.602
	2	91.700	91.964
	3	87.325	88.460
	4	90.375	89.338
	5	90.825	92.249
	6	89.325	90.346
	7	89.825	90.285
	8	90.050	90.438
	9	98.600	91.192
	10	101.600	94.334
	11	102.400	93.306
	12	95.900	92.616
	13	99.950	92.544
	14	100.625	92.303
	15	95.300	89.979
	16	97.525	94.114
	17	76.725	86.127
	18	80.400	87.132
	19	80.175	86.665
	20	90.750	89.907
	21	77.875	85.803
	22	81.525	87.338
	23	82.950	87.674
	24	80.225	86.520
	25	81.225	87.830
	26	81.525	89.456
	27	99.050	91.991
Summary Statistics	Mean	89.817	89.982
	Standard Deviation	8.005	2.516
	Correlation	0.923	
	Regression Line	$Lab_{\text{New}}/Lab_{\text{Old}} = 63.94 + 0.29 \times Lab_{\text{Old}}$	

where $\hat{\alpha}$ and $\hat{\beta}$ are the least-squares estimates, and

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(\hat{\alpha} + \hat{\beta}Y_{1i} - Y_{2i})^2$$

Thus, A_ρ can be estimated by

$$\tilde{\mathrm{A}}_\rho = \frac{4S_1\sqrt{\hat{\beta}_2 S_1^2 + \hat{\sigma}^2} - \hat{\rho}(S_1^2 + \hat{\beta}^2 S_1^2 + \hat{\sigma}^2)}{(2-\hat{\rho})(S_1^2 + \hat{\beta}^2 S_1^2 + \hat{\sigma}^2) + (\bar{\mathrm{Y}}_2 - \bar{\mathrm{Y}}_1)^2}$$

This approach considers the measurement error of the 'gold' standard. A further research study will be conducted to see which approach performs better based on the mean squared error.

ACKNOWLEDGEMENTS

The author thanks the editor, Dr. Christy Chuang-Stein, the two referees, and Dr. Seth Clark for their constructive comments which improved the presentation of this paper.

REFERENCES

1. Vonesh EF, Chinchilli VM. *Linear and nonlinear models for the analysis of repeated measurements*. Marcel Dekker: New York, 1997.
2. MacLennan RN. Interrater reliability with SPSS for Windows 5.0. *American Statistician* 1993; **47**:292–296.

3. Guerrero, J-L, Roshwalb A. A measure of agreement for use in assessing record quality. *ASA Proceedings of the Section on Quality and Productivity* 1991;82–87.
4. Hofstee WKB, Zegers FE. Idiographic Correlation: Modeling Judgments of Agreement between school grades. *Tijdschrift Voor Onderwijs Research* 1991; **16**:331–336.
5. Krippendorff K. Association, agreement, and equity. *Quality and Quantity* 1987; **21**:109–123.
6. Tanner MA, Young MA. Modeling agreement among raters. *Journal of the American Statistical Association* 1985; **80**:175–180.
7. Lin L. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; **45**:255–268.
8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **2**:307–310.
9. Ludbrook J. Comparing methods of measurement. *Clinical and Experimental Pharmacology and Physiology* 1997; **24**:193–203.
10. Lin LI-K, Torbeck LD. Coefficient of accuracy and concordance correlation coefficient: new statistics for methods comparison. *PDA Journal of Pharmaceutical Science & Technology* 1998; **52**(2):55–59.
11. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Statistics in Medicine* 1994; **13**:2465–2476.
12. Nickerson CAE. A note on 'A concordance correlation coefficient to evaluate reproducibility'. *Biometrics* 1997; **53**:1503–1507.
13. Liao J, Lewis J. A note on concordance correlation coefficient. *PDA Journal of Pharmaceutical Science & Technology* 2000; **54**(1):23–26.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**:159–174.