

Chapter 2

Continuous Data

We will now introduce new approaches that have evolved for measuring agreement since 1988. Some of these new approaches were summarized, studied, and compared by Lin, Hedayat, Sinha, and Yang (2002). Here, we include the necessary proofs that were left out of that article. In addition, we include assorted examples to demonstrate agreement techniques. We begin with the most basic model, in which paired observations (Y and X) are collected.

2.1 Basic Model

When target values are random, the joint distribution of Y and X is assumed to have a bivariate distribution with finite second moments with means μ_y, μ_x , variances σ_y^2, σ_x^2 and covariance σ_{yx} . When target values are fixed, $Y_i|X_i, i = 1, \dots, n$, are assumed to be observations in a random sample from the basic regression model $Y = \beta_0 + \beta_1 X + e_Y$. Here, e_Y is the residual error with mean 0 and variance σ_e^2 .

2.2 Absolute Indices

2.2.1 Mean Squared Deviation

Mean squared deviation (MSD) evaluates an aggregated deviation from the identity line, $\text{MSD} = E(Y - X)^2$. It can be expressed as

$$\varepsilon^2 = (\mu_y - \mu_x)^2 + \sigma_y^2 + \sigma_x^2 - 2\sigma_{yx}, \quad (2.1)$$

when target values are random, or

$$\varepsilon_{|X}^2 = (\mu_{y|X} - \bar{X})^2 + s_x^2(1 - \beta_1)^2 + \sigma_e^2, \quad (2.2)$$

when target values are fixed, where \bar{X} and s_x^2 are the sample mean and variance of X .

Estimated by sample counterparts (e^2 or $e_{|X}^2$) with a log transformation, $W = \ln(e^2)$ or $W_{|X} = \ln(e_{|X}^2)$ has an asymptotic normal distribution with mean $\ln(\varepsilon^2)$ or $\ln(\varepsilon_{|X}^2)$ and variance

$$\sigma_W^2 = \frac{2}{n-2} \left[1 - \frac{(\mu_y - \mu_x)^4}{\varepsilon^4} \right], \quad (2.3)$$

when target values are random, or

$$\sigma_{W|X}^2 = \frac{2}{n-2} \left[1 - \frac{(\varepsilon_{|X}^2 - \sigma_e^2)^2}{\varepsilon_{|X}^4} \right], \quad (2.4)$$

when target values are fixed.

The proof of (2.3) can be found in Section 2.9.2. The proof of (2.4) can be found in Section 2.10.2. For statistical inference, refer to Section 2.4 for the information regarding sample counterparts, and to Sections 2.9 and 2.10 for the proofs of asymptotic normality in this chapter.

The MSD is not an easy index to interpret. The following methods will put some meaningful interpretation on this basic MSD index.

2.2.2 Total Deviation Index

To examine the agreement from a different perspective, a measure that captures a large proportion (π_0) of data within a boundary (δ_0) from target values was considered by Lin, Hedayat, Sinha, and Yang (2002). For example, we may want to capture at least 90% of individual observations that are within 10% of their target values. We would compute total deviation index (TDI) for the given coverage probability (CP) criterion of 0.9 to see whether this TDI is less than 10%, or compute coverage probability (CP) for the given TDI criterion of 10% to see whether this CP is more than 0.9.

Assume that $D = Y - X$ has a normal distribution with mean $\mu_d = \mu_y - \mu_x$ and variance $\sigma_d^2 = \sigma_y^2 + \sigma_x^2 - 2\sigma_{xy}$. We may find π for a given δ_0 criterion, CP_{δ_0} , which is

$$\pi_{\delta_0} = P(D^2 < \delta_0^2) = \chi^2 \left(\delta_0^2, 1, \frac{\mu_d^2}{\sigma_d^2} \right), \quad (2.5)$$

where $\chi^2(\cdot)$ is the cumulative noncentral chi-square distribution up to δ_0^2 , with one degree of freedom and noncentrality parameter $\frac{\mu_d^2}{\sigma_d^2}$. This measure will be presented shortly.

We may also find δ for a given π_0 criterion, TDI_{π_0} , which is

$$\delta_{\pi_0} = \sqrt{(\chi^2)^{-1} \left(\pi_0, 1, \frac{\mu_d^2}{\sigma_d^2} \right)}, \quad (2.6)$$

where $(\chi^2)^{-1}$ is the inverse function of $\chi^2(\cdot)$. Since the estimate of this index has intractable asymptotic properties, Lin (2000) and Lin, Hedayat, Sinha, and Yang (2002) have suggested the following TDI_{π_0} approximation:

$$\delta_{\pi_0 \sim}^2 \doteq (\chi^2)^{-1}(\pi_0, 1) \varepsilon^2, \quad (2.7)$$

or

$$\delta_{\pi_0 \sim} \doteq \Phi^{-1} \left(1 - \frac{1 - \pi_0}{2} \right) |\varepsilon|, \quad (2.8)$$

when X is random, or

$$\delta_{\pi_0 \sim |X} \doteq \Phi^{-1} \left(1 - \frac{1 - \pi_0}{2} \right) |\varepsilon|_X, \quad (2.9)$$

when X is fixed.

The approximation is satisfactory (Lin 2000) when:

1. $\pi_0 = 0.75$ and $\Delta \leq 1/2$,
2. $\pi_0 = 0.8$ and $\Delta \leq 8$,
3. $\pi_0 = 0.85$ and $\Delta \leq 2$,
4. $\pi_0 = 0.9$ and $\Delta \leq 1$,
5. $\pi_0 = 0.95$ and $\Delta \leq 1/2$.

The quantity $\Delta = \frac{\mu_d^2}{\sigma_d^2}$ is called the *relative bias squared* (RBS). The interpretation of this approximated TDI is that approximately $100\pi_0\%$ of observations are within $\delta_{\pi_0 \sim}$ of the target values. $\text{TDI}_{\pi_0}^2$ is proportional to MSD, and therefore we may perform an inference based on the asymptotic normality of $W = \ln(e^2)$, where e^2 is the sample counterpart of MSD when X is random, or $W|_X = \ln(e^2|_X)$ when X is fixed. This simplified method will become very useful when we deal with the more general case to be introduced in Chapter 5.

The idea of using such an approximation of TDI was motivated by Holder and Hsuan (1993). They proposed a moment-based criterion for assessing individual bioequivalence. They showed that in a slightly different fashion, $\delta_{\pi_0}^2$, or the squared function of (2.6), has an upper bound $\delta_{\pi_0+}^2 = c_{\pi_0} \varepsilon^2$, where c_{π_0} is a constant not depending on μ_d and σ_d . Therefore, δ_{π_0+} conservatively captures at least $100\pi_0\%$ of observations within the boundary from target values of a reference compound. Holder and Hsuan (1993) used a numerical algorithm for the determination of c_{π_0} under some parametric and nonparametric distribution of $D = Y - X$.

However, the asymptotic distribution property of this estimate has not been established. Lin (2000) made a comparison between this statistic and the TDI given in (2.7). When $\pi_0 = 0.9$, $\delta_{\pi_0\sim}^2$ and $\delta_{\pi_0+}^2$ are identical under the normality assumption. Using $\text{TDI}_{0.9}$ is almost exact when $\frac{\mu_d}{\sigma_d} < 1$, and would become conservative otherwise. Using $\text{TDI}_{0.8}$ is most robust, since it can tolerate an RBS value as high as 8.0.

A TDI is similar in concept to a tolerance limit. The difference is that a tolerance limit captures individual deviations from their own mean, while a TDI captures individual deviations from their target values, for a high proportion (say, 90%), and with a high degree of confidence (say, 95%) when the upper confidence limit of TDI is used.

2.2.3 Coverage Probability

We now consider finding π for a given δ_0 criterion. This is

$$\pi_{\delta_0} = P(D^2 < \delta_0^2) = \chi^2 \left(\delta_0^2, 1, \frac{\mu_d^2}{\sigma_d^2} \right), \quad (2.10)$$

when target values are random, or

$$\pi_{\delta_0|X} = \frac{1}{n} \sum_{i=1}^n \pi_{\delta_0(i)}, \quad (2.11)$$

where

$$\pi_{\delta_0(i)} = \chi^2 \left[\frac{\delta_0^2}{\sigma_e^2}, 1, \left(\frac{\beta_0 + (\beta_1 - 1)X_i}{\sigma_e} \right)^2 \right], \quad (2.12)$$

when target values are fixed. The estimate of coverage probability using sample counterparts (p_{δ_0} or $p_{\delta_0|X}$) by the logit transformation, $T = \ln \frac{p_{\delta_0}}{1-p_{\delta_0}}$ or $T|X = \ln \frac{p_{\delta_0|X}}{1-p_{\delta_0|X}}$, has an asymptotic normal distribution with mean $E(p_{\delta_0}) = \ln \frac{\pi_{\delta_0}}{1-\pi_{\delta_0}}$ or $E(p_{\delta_0|X}) = \ln \frac{\pi_{\delta_0|X}}{1-\pi_{\delta_0|X}}$, and variance

$$\sigma_T^2 = \frac{0.5 [\delta_{+\mu\sigma} \phi(-\delta_{+\mu\sigma}) + \delta_{-\mu\sigma} \phi(\delta_{-\mu\sigma})]^2 + [\phi(-\delta_{+\mu\sigma}) - \phi(\delta_{-\mu\sigma})]^2}{(n-3)(1-\pi_{\delta_0})^2 \pi_{\delta_0}^2}, \quad (2.13)$$

when X is random, or

$$\sigma_{T|X}^2 = \frac{\left[\frac{C_0^2}{n^2} + \frac{(C_0\bar{X}-C_1)^2}{n^2 s_X^2} + \frac{C_2^2}{2n^2} \right]}{(n-3)(1-\pi_{\delta_0})^2 \pi_{\delta_0}^2}, \quad (2.14)$$

when X is fixed, where

$$\delta_{+\mu\sigma} = \frac{\delta_0 + \mu_d}{\sigma_d}, \quad (2.15)$$

$$\delta_{-\mu\sigma} = \frac{\delta_0 - \mu_d}{\sigma_d}, \quad (2.16)$$

$$\delta_{+\mu\beta\sigma_i} = \frac{\delta_0 + \beta_0 + (\beta_1 - 1)X_i}{\sigma_e}, \quad (2.17)$$

$$\delta_{-\mu\beta\sigma_i} = \frac{\delta_0 - \beta_0 - (\beta_1 - 1)X_i}{\sigma_e}, \quad (2.18)$$

$$C_0 = \sum_{i=1}^n [\phi(-\delta_{+\mu\beta\sigma_i}) - \phi(\delta_{-\mu\beta\sigma_i})], \quad (2.19)$$

$$C_1 = \sum_{i=1}^n [\phi(-\delta_{+\mu\beta\sigma_i}) - \phi(\delta_{-\mu\beta\sigma_i})] X_i, \quad (2.20)$$

$$C_2 = \sum_{i=1}^n [\delta_{+\mu\beta\sigma_i} \phi(-\delta_{+\mu\beta\sigma_i}) + \delta_{-\mu\beta\sigma_i} \phi(\delta_{-\mu\beta\sigma_i})], \quad (2.21)$$

and $\phi(\cdot)$ is the standard normal distribution function.

The proof of (2.13) can be found in Section 2.9.3. The proof of (2.14) can be found in Section 2.10.3.

We can also use the simplified method of

$$\pi_{\delta_0 \sim} \doteq \chi^2 \left(\frac{\delta_0^2}{\varepsilon^2}, 1 \right), \quad (2.22)$$

which will become very useful when we deal with the more general case to be introduced in Chapter 5.

2.3 Relative Indices

2.3.1 Intraclass Correlation Coefficient

Pearson (1899, 1901) developed the intraclass correlation coefficient (ICC) as a way to estimate various aspects of fraternal resemblance. Given pairs of brothers, one might be interested not in the correlation in height between the older and the younger brother, or the taller and the shorter brother, but simply between brothers in general. In this case, the heights of the two brothers are logically interchangeable. Pearson suggested that this correlation could be estimated by entering the height

measurements for each pair of brothers, (x, y) , twice into the computation of the usual product-moment correlation coefficient ρ , once in the order (x, y) and once in the order (y, x) . If there are more than two brothers in each set, each possible pair of measurements is entered twice into the computation. Thus, the number of entries in the correlation for this set is $n(n - 1)$, where n is the number of brothers in the data set.

Harris (1913) developed a simple formula for intraclass correlation as a function of (a) the variance of the means of each set of measurements around the overall mean and (b) the variance of the total set of measurements.

Fisher (1925) observed that variance measurements could be partitioned into two components. The first component is the between-sample variance after removing the residual variance, which Fisher called A. The second component is the residual variance or within-sample variance, which Fisher called B. Thus the population intraclass correlation can be expressed as

$$\rho_I = \frac{A}{A + B}. \quad (2.23)$$

Fisher (1925) noted that the ICC could be estimated using mean squares from an analysis of variance (ANOVA). We will revisit ICC in Chapter 3, where we will show its association with kappa, weighted kappa, and the concordance correlation coefficient (CCC) presented below. We will also revisit, in Chapter 5, the general form of the ICC for agreement, precision, and accuracy coefficients.

2.3.2 Concordance Correlation Coefficient

The MSD can be standardized such that: 1 indicates that each pair of readings is in perfect agreement in the population (for example, 1, 1; 2, 2; 3, 3; 4, 4; 5, 5), while 0 indicates no correlation, and -1 means that each pair of readings is in perfect reversed agreement in the population (for example, 5, 1; 4, 2; 3, 3; 2, 4; 1, 5). Lin (1989) introduced one such standardization of MSD, called CCC, which is defined as

$$\rho_c = 1 - \frac{\varepsilon^2}{\varepsilon^2_{|\rho=0}} \quad (2.24)$$

$$\begin{aligned} &= 1 - \frac{\varepsilon^2}{\sigma_y^2 + \sigma_x^2 + (\mu_y - \mu_x)^2} \\ &= \frac{2\sigma_{yx}}{\sigma_y^2 + \sigma_x^2 + (\mu_y - \mu_x)^2}, \\ &= \frac{2\rho\sigma_x\sigma_y}{\sigma_y^2 + \sigma_x^2 + (\mu_y - \mu_x)^2} \end{aligned} \quad (2.25)$$

when X is random, and

$$\rho_{c|X} = \frac{2\beta_1 s_x^2}{\sigma_y^2 + s_x^2 + (\mu_y - \bar{X})^2}, \quad (2.26)$$

when X is fixed. The CCC is closely related to the intraclass correlation and has a meaningful geometrical interpretation. It is inversely related to the mean square of the ratio of the within-sample total deviation (ε^2) and the total deviation ($\varepsilon_{|\rho=0}^2$). For example, if the within-sample total deviation is 10%, 32%, or 45% of the total deviation, then the CCC is $0.99 = (1 - 0.1^2)$, $0.90 = (1 - 0.32^2)$, or $0.80 = (1 - 0.45^2)$, respectively. In Chapter 3, we will show that for ordinal categorical data, CCC degenerates into the weighted kappa suggested by Cohen (1968).

Section 1.1 defined accuracy and precision in the one-dimensional situation. According to the two-dimensional model of Section 2.1, the between-sample variation is typically inherited or is a result of the design of the sampling process, and is usually unrelated to within-sample precision of an assay. Therefore, we consider the difference in between-sample variance as a systematic bias, and it is included in the inaccuracy. A sample mean and sample variance define a marginal distribution in most of the commonly used distributions.

2.3.2.1 Accuracy Coefficient

The accuracy coefficient measures the closeness of the marginal distributions of Y and X , where 1 signifies equal means and variances, and 0 indicates that the absolute difference in means and/or variance approach infinity. The accuracy coefficient can be broken down into measures of location and/or scale shifts, where the location shift is $\nu = \frac{\mu_y - \mu_x}{\sqrt{\sigma_y \sigma_x}}$, and the scale shift is $\varpi = \frac{\sigma_y}{\sigma_x}$ or $\frac{\sigma_x}{\sigma_y}$. Here, the accuracy coefficient is defined as

$$\chi_a = \frac{2}{\varpi + 1/\varpi + \nu^2}, \quad (2.27)$$

when X is random. We can replace σ_x^2 by s_x^2 , and μ_x by \bar{X} in (2.27) when X is fixed.

2.3.2.2 Precision Coefficient

The precision coefficient is the Pearson correlation coefficient (ρ) between Y and X , where

$$\rho = \frac{\sigma_{yx}}{\sigma_y \sigma_x}. \quad (2.28)$$

Here ρ^2 has the same scale as the accuracy coefficient, from 0 (no agreement) to 1 (perfect agreement). It is evident from (2.25), (2.26), and (2.27) that the CCC is the product of the precision and accuracy coefficients when target values are random or fixed.

2.3.2.3 Statistical Inference on CCC and the Accuracy and Precision Coefficients

The estimate of CCC (r_c or $r_{c|X}$) using sample counterparts by the Z transformation, $Z = \frac{1}{2} \ln \frac{1+r_c}{1-r_c}$ or $Z_{|X} = \frac{1}{2} \ln \frac{1+r_{c|X}}{1-r_{c|X}}$, has an asymptotic normal distribution with mean $\frac{1}{2} \ln \frac{1+\rho_c}{1-\rho_c}$ or $\frac{1}{2} \ln \frac{1+\rho_{c|X}}{1-\rho_{c|X}}$, and variance

$$\sigma_Z^2 = \frac{1}{(n-2)} \left[\frac{(1-\rho^2)\rho_c^2}{(1-\rho_c^2)\rho^2} + \frac{2\rho_c^3(1-\rho_c)v^2}{\rho(1-\rho_c^2)^2} - \frac{\rho_c^4 v^4}{2\rho^2(1-\rho_c^2)^2} \right] \quad (2.29)$$

or

$$\sigma_{Z|X}^2 = \frac{(1-\rho^2)\rho_c^2}{[(n-2)(1-\rho_c^2)^2\rho^2]} \left[\varpi v^2 \rho_c^2 + (1-\rho_c \rho \varpi)^2 + \frac{\varpi^2 \rho_c^2 (1-\rho^2)}{2} \right], \quad (2.30)$$

when X is random or fixed, respectively.

The estimate of the accuracy coefficient (c_a or $c_{a|X}$) using sample counterparts by the logit transformation, $L = \ln \frac{c_a}{1-c_a}$ or $L_{|X} = \ln \frac{c_{a|X}}{1-c_{a|X}}$, has an asymptotic normal distribution with mean $\ln \frac{\chi_a}{1-\chi_a}$ or $\ln \frac{\chi_{a|X}}{1-\chi_{a|X}}$, and variance

$$\sigma_L^2 = \frac{\chi_a^2 v^2 (\varpi + 1/\varpi - 2\rho) + \frac{1}{2} \chi_a^2 (\varpi^2 + 1/\varpi^2 + 2\rho^2) + (1 + \rho^2)(\chi_a v^2 - 1)}{(n-2)(1-\chi_a)^2} \quad (2.31)$$

or

$$\sigma_{L|X}^2 = \frac{v^2 \varpi \chi_a^2 (1-\rho^2) + \frac{1}{2} (1-\varpi \chi_a)^2 (1-\rho^4)}{(n-2)(1-\chi_a)^2}. \quad (2.32)$$

The estimate of the precision coefficient (r or $r_{|X}$) by the Z -transformation has an asymptotic normal distribution with mean $\frac{1}{2} \ln \frac{(1+\rho)}{(1-\rho)}$ or $\frac{1}{2} \ln \frac{(1+\rho_{|X})}{(1-\rho_{|X})}$, and variance $\frac{1}{n-3}$ when target values are random, or $\frac{1-\rho^2/2}{n-3}$ when target values are fixed. The proof of (2.29) and (2.30) can be found in Sections 2.9.1 and 2.10.1. The proof of (2.31) and (2.32) can be found in Sections 2.9.4 and 2.10.4.

2.4 Sample Counterparts

For the purpose of statistical inference, the parameters discussed above can be replaced with their sample estimates that are consistent estimators, such as the moments estimates.

The sample counterparts for μ_y , μ_x , σ_y^2 , σ_x^2 , σ_{yx} , β_1 , and ρ are

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad (2.33)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad (2.34)$$

$$s_y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2, \quad (2.35)$$

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad (2.36)$$

$$s_{yx}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X}), \quad (2.37)$$

$$b_1 = \frac{s_{yx}}{s_x^2}, \quad (2.38)$$

and

$$r = \frac{s_{yx}}{s_y s_x}. \quad (2.39)$$

We could certainly use $\frac{1}{n-1}$ rather than $\frac{1}{n}$ in the above variance and covariance estimates. However, we use $\frac{1}{n}$ to bound the CCC estimate by 1.0.

In estimating the MSD, we use the sum of squared difference divided by $n - 1$. For (2.17) and (2.18) when X is fixed, we use

$$s_e^2 = \frac{n}{(n-3)}(1-r^2)s_y^2. \quad (2.40)$$

For less bias in estimating the RBS, and in estimating σ_d^2 in (2.15) and (2.16), we use

$$s_d^2 = \frac{n}{(n-3)}(s_x^2 + s_y^2 - 2s_{xy}). \quad (2.41)$$

The use of $n - 2$ or $n - 3$ instead of n in the denominators of the above variance equations is for the small-sample-size bias correction based on the simulation studies in Lin, Hedayat, Sinha, and Yang (2002). The use of sample-size bias correction is not important when the sample size is large.

For the purpose of performing statistical inference for each index, we should compute the confidence limit (lower limit for CCC, precision coefficient, accuracy coefficient, CP, and upper limit for TDI) based on its respective transformation, then perform antitransformation to the limit. We will declare that the assay agreement is acceptable when the limit is better than the prespecified criterion. The use of transformed estimates can speed up the approach to normality. Moreover, a transformation could bound the confidence interval to its respective parameter range, say, -1 to 1 for CCC and precision coefficient, 0 to 1 for the accuracy coefficient and CP, and 0 to infinity for MSD.

Throughout this book, once the asymptotic normality of an estimated index has been defined, statistical inference can be established through confidence limit(s). Let $\hat{\lambda}$ be the estimate of an agreement index and $\sigma_{\hat{\lambda}}^2$ its variance. Then the one-sided upper or lower confidence limit becomes

$$\hat{\lambda} + \Phi^{-1}(1 - \alpha)\hat{\sigma}_{\hat{\lambda}} \quad \text{or} \quad \hat{\lambda} - \Phi^{-1}(1 - \alpha)\hat{\sigma}_{\hat{\lambda}},$$

where $\hat{\sigma}_{\hat{\lambda}}$ is the estimate of the square root of variance, $\sigma_{\hat{\lambda}}$, using sample counterparts. When the sample size is small, say less than 30, we can also use the cutoff value of the cumulative central t -distribution instead of the standard cumulative normal distribution to form the statistical inference.

2.5 Proportional Error Case

When Y and X are positively valued variables and the standard deviations of Y are proportional to either Y or X , it is assumed that $\ln(Y)$ and $\ln(X)$ have a bivariate normal distribution. Let $100\theta\%$ be the percent change in Y and X . Then

$$\pi_{\theta} = P \left[\frac{1}{(1 + \theta)} < \frac{Y}{X} < (1 + \theta) \right] = P [|\ln(Y) - \ln(X)| < \ln(1 + \theta)]. \quad (2.42)$$

Let $D = \ln(Y) - \ln(X)$ and $\delta_{\pi_0} = \ln(1 + \theta_{\pi_0})$. Then $\theta_{\pi_0} = 100[\exp(\delta_{\pi_0}) - 1]\%$. This $100\theta_{\pi_0}\%$ is denoted by $\text{TDI}\%_{\pi_0}$.

In the case of proportional errors, all of the above unscaled and scaled agreement indices should be computed from the log transformed data. In practice, we have encountered the proportional error case more frequently than the constant error case.

2.6 Summary of Simulation Results

A comprehensive simulation was conducted in Lin, Hedayat, Sinha, and Yang (2002) to study the small-sample properties of the transformed estimates of CCC, precision coefficient, accuracy coefficient, TDI, and CP. The results showed

excellent agreement with the theoretical values from normal samples even when $n = 15$. However, these estimates are not expected to be robust against outliers or large deviations from normality or log-normality. The robustness issues of the CCC have been addressed in King and Chinchilli (2001a, 2001b), using M-estimation or using a power function of the absolute value of D to compute the CCC.

2.7 Asymptotic Power and Sample Size

In assessing agreement, the null and alternative hypotheses should be reversed. The conventional rejection region actually is the region of declaring agreement (one-sided). Asymptotic power and sample size calculation should proceed by the above principle. These powers of CCC, TDI, and CP were compared in Lin, Hedayat, Sinha, and Yang (2002). The results showed that the TDI and CP estimates have similar power, and are superior to CCC, but they are valid only under the normality assumption. Therefore, for inference, TDI and CP are superior to CCC. However, the CCC and precision and accuracy coefficients remain very useful and informative tools, as is evident from the following examples. In Chapter 4, we will discuss the sample size subject in greater detail.

2.8 Examples

2.8.1 Example 1: Methods Comparison

This example was presented in Lin, Hedayat, Sinha, and Yang (2002). DCLHb is a treatment solution containing oxygen-carrying hemoglobin. The DCLHb level in a patient's serum is routinely measured by the Sigma method. The simpler HemoCue method was modified to reproduce the DCLHb values of the Sigma method. Serum samples from 299 patients over a 50–2,000 mg/dL range were collected. The DCLHb values of each sample were measured by both methods twice, and the averages of the duplicate values were evaluated. The client required with 95% confidence that the within-sample total deviation be less than 15% of the total deviation. This means that the allowable CCC was $1 - 0.15^2 = 0.9775$. The client also needed with 95% confidence that at least 90% of the HemoCue observations be within 150 mg/dL of the targeted Sigma values. This means that the allowable $TDI_{0.9}$ was 150 mg/dL, or that the allowable CP_{150} was 0.9.

The results are presented in Fig. 2.1 and Table 2.1. The plot indicates that the within-sample error is relatively constant across the clinical range. The plot also indicates that the HemoCue accuracy is excellent and that the precision is adequate. The CCC estimate is 0.987, which means that the within-sample total deviation is about 11.6% of the total deviation. The CCC one-sided lower confidence limit is

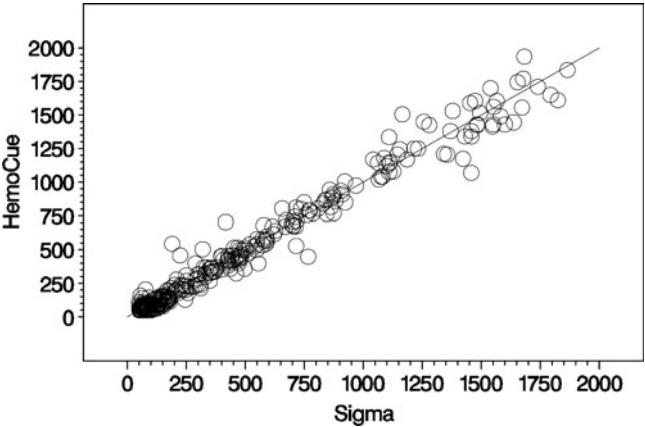


Fig. 2.1 HemoCue and Sigma readings on measuring DCLHb

Table 2.1 Agreement statistics for HemoCue and Sigma readings on measuring DCLHb

Statistics	CCC	Precision coefficient	Accuracy coefficient	$TDI_{0.9}$	CP_{150}	RBS
Estimate	0.9866	0.9867	0.9999	127.5	0.9463	0.00
95% Conf. limit	0.9838	0.9839	0.9989	136.4	0.9276	–
Allowance	0.9775	–	–	150.0	0.9000	–

“–” means “not applicable”

0.984, which is greater than 0.9775. The precision coefficient estimate is 0.987 with a one-sided lower confidence limit of 0.984. The accuracy coefficient estimate is 0.9999 with a one-sided lower confidence limit of 0.9989. The $TDI_{0.9}$ estimate is 127.5 mg/dL, which means that 90% of HemoCue observations are within 127.5 mg/dL of their target values. The one-sided upper confidence limit for $TDI_{0.9}$ is 136.4 mg/dL, which is less than 150 mg/dL. Finally, the CP_{150} estimate is 0.946, which means that 94.6% of HemoCue observations are within 150 mg/dL of their target values. The one-sided lower confidence limit for CP_{150} is 0.928, which is greater than 0.9. Therefore, the agreement between HemoCue and Sigma is acceptable with excellent accuracy and adequate precision. The relative bias squared is estimated to be near zero, indicating that the approximation of TDI should be excellent.

2.8.2 Example 2: Assay Validation

This example was presented in Lin, Hedayat, Sinha, and Yang (2002). FVIII is a clotting agent in plasma. The FVIII assay uses a marker with varying dilutions of known FVIII activities to form a standard curve. The assay started at 1:5 or 1:10

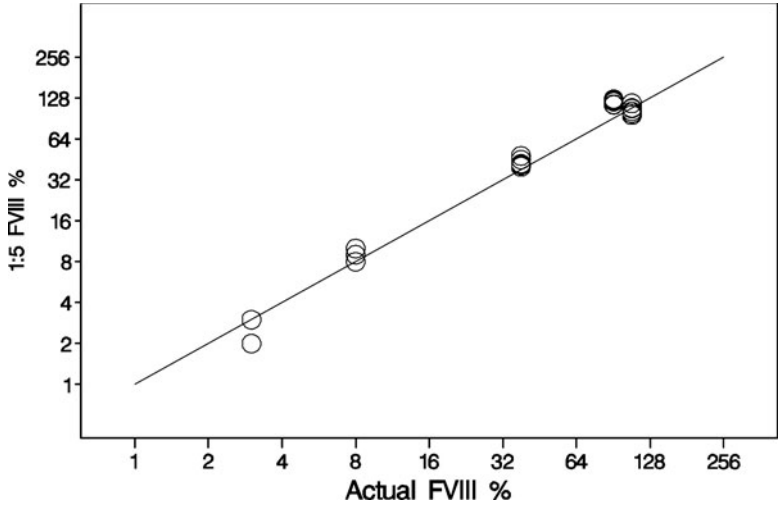


Fig. 2.2 Observed FVIII assay results versus targeted values started at 1:5

serial dilutions were prepared until they reached the target values. Target values were fixed at 3%, 8%, 38%, 91%, and 108%. Six samples were assayed per target value. The error was expected to be proportional mainly due to dilutions. The client needed with 95% confidence that the within-sample total deviation be less than 15% of the total deviation. This means that the allowable CCC was $1 - 0.15^2 = 0.9775$. The client also needed with 95% confidence that 80% of FVIII observations be within 50% of target values (note that this is the percentage of the measuring unit, which is also a percentage). This means that the allowable $TDI_{0.8}$ was 50%, or that the allowable $CP_{50\%}$ was 0.8.

Figures 2.2 and 2.3 present the results started at 1:5 and at 1:10 serial dilutions for these plots of observed FVIII assay results versus targeted values in log₂ scale. Note that there are overlying observations in the plots. Specifically, in Fig. 2.2, four replicate readings of 3% and duplicate readings of 2% are observed at the target value of 3%, and circles at the target value of 8% represent duplicate readings of 8%, 9%, and 10%. Duplicate readings of 45% are observed at target values of 38%. Also note that in Fig. 2.3, four replicate readings of 5% and duplicate readings of 4% are observed at the target value of 3%. Three replicate readings of 11% and duplicate readings of 12% are observed at the target value of 8%. Duplicate readings of 49% are observed at target values of 38%. Duplicate readings of 124% are observed at target values of 91%. The plots indicate that the within-sample error is relatively constant across the target values in log scale. The precision is good for both assays started at 1:5 and at 1:10 serial dilutions, but the accuracy is not as good for the assay started at 1:10 serial dilutions.

Tables 2.2 and 2.3 present the agreement statistics started at 1:5 and 1:10 serial dilutions. For the assay started at 1:5 serial dilutions, the CCC is estimated to be 0.992, which means that the within-sample total deviation is about 9.1% of the

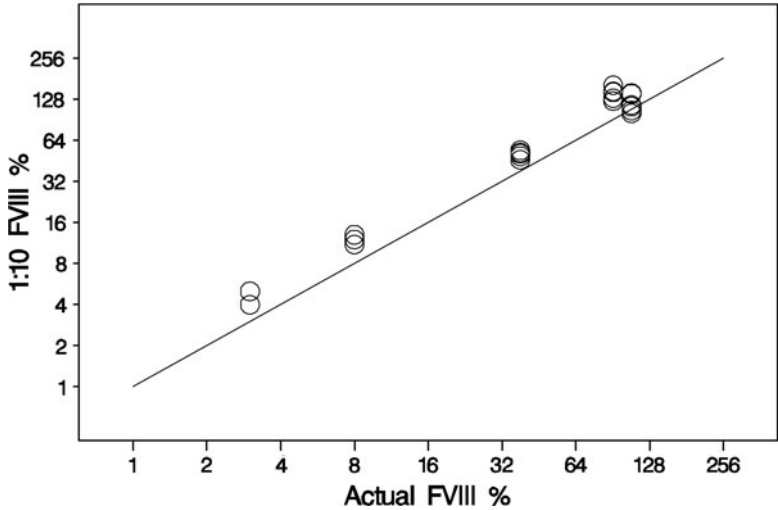


Fig. 2.3 Observed FVIII assay results versus targeted values started at 1:10

Table 2.2 FVIII assay results started at 1 : 5

Statistics	CCC	Precision coefficient	Accuracy coefficient	TDI% _{0.8}	CP _{50%}	RBS
Estimate	0.9917	0.9942	0.9975	27.35	0.9653	0.12
95% Conf. limit	0.9875	0.9908	0.9935	35.01	0.8921	–
Allowance	0.9775	–	–	50.0	0.8000	–

“–” means “not applicable”

Table 2.3 FVIII assay results started at 1:10

Statistics	CCC	Precision coefficient	Accuracy coefficient	TDI% _{0.8}	CP _{50%}	RBS
Estimate	0.9669	0.9947	0.9721	58.95	0.7016	3.75
95% Conf. limit	0.9584	0.9917	0.9638	69.01	0.5898	–
Allowance	0.9775	–	–	50.0	0.8000	–

“–” means “not applicable”

total deviation. The one-sided lower confidence limit is 0.987, which is greater than 0.9775. The precision coefficient is estimated to be 0.994 with a one-sided lower confidence limit of 0.991. The accuracy coefficient is estimated to be 0.998 with a one-sided lower confidence limit of 0.994. $TDI_{0.8}$ is estimated to be 27.3%, which means that 80% of observations are within 27.3% change from target values (percentage of percentage values). The one-sided upper confidence limit is 35.0%, which is less than 50%. Finally, $CP_{50\%}$ is estimated to be 0.965, which means that 96.5% of observations are within 50% change from target values. The one-sided lower confidence limit is 0.892, which is greater than 0.8. The agreement between

the FVIII assay and the actual concentration is acceptable with good precision and accuracy. The relative bias squared is estimated to be 0.12, so that the approximation of TDI should be excellent.

For the assay started at 1:10 serial dilutions, the CCC is estimated to be 0.967, which means that the within-sample total deviation is about 18.2% of the total deviation. The one-sided lower confidence limit is 0.958, which is less than 0.9775. The precision coefficient is estimated to be 0.995 with a one-sided lower confidence limit of 0.992. The accuracy coefficient is estimated to be 0.972 with a one-sided lower confidence limit of 0.964. $TDI_{\%0.8}$ is estimated to be 58.9%, which means that 80% of observations are within 58.9% change from target percentage values. The one-sided upper confidence limit is 69.0%, which is greater than 50%. Finally, $CP_{50\%}$ is estimated to be 0.702, which means that 70.2% of observations are within 50% change from target values. The one-sided lower confidence limit is 0.590, which is less than 0.8. The agreement between the FVIII assay and actual concentration had good precision but is not acceptable due to mediocre accuracy. The relative bias squared is estimated to be 3.75, which is less than 8.0, so that the approximation of TDI should be acceptable.

2.8.3 Example 3: Assay Validation

This example was presented in Lin and Torbeck (1998). A study to validate an amino acid analysis test method was conducted. Solutions were prepared at approximately 90%, 100%, and 110% of label concentration of the amino acids, each containing nine determinations (observed values). Target values were determined based on their average molecular weights, which were much more precise and accurate but were still measured with random error. For each test method we compute the estimates of CP, TDI, CCC, precision coefficient, accuracy coefficient, and their confidence limits. It is debatable whether we should treat the target values as random or fixed because they were average values. We therefore take the more conservative approach by treating target values as random, which yields the same estimates of agreement statistics but with a larger respective standard error for each estimate.

The observed and target values were expressed as a percentage of label concentration. Using estimates of the CCC components, coefficients of accuracy (c_a) and precision (r), four out of 30 amino acids were chosen for illustration, each representing an example of distinctive precise and/or accurate situations. These four amino acids and their label concentrations were glycine (1 g/L), ornithine (6.4 g/L), L-threonine acid (3.6 g/L), and L-methionine (2 g/L).

The range of these data was approximately 20% (90%–110%) of label concentration. The client needed with 95% confidence that at least 80% of observations be within 3% of target values. This means that the 95% upper limit of $TDI_{0.8}$ must be less than 3, or that the 95% lower limit of CP_3 must be greater than 0.8. Note that the measurement unit is in percentage, and the error structure was assumed constant across the data range. The client did not specify a criterion for the CCC.

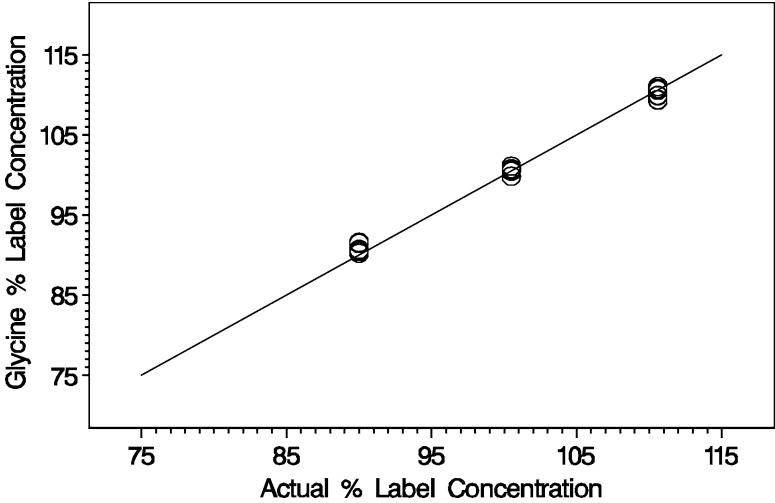


Fig. 2.4 Observed measures versus target values of glycine

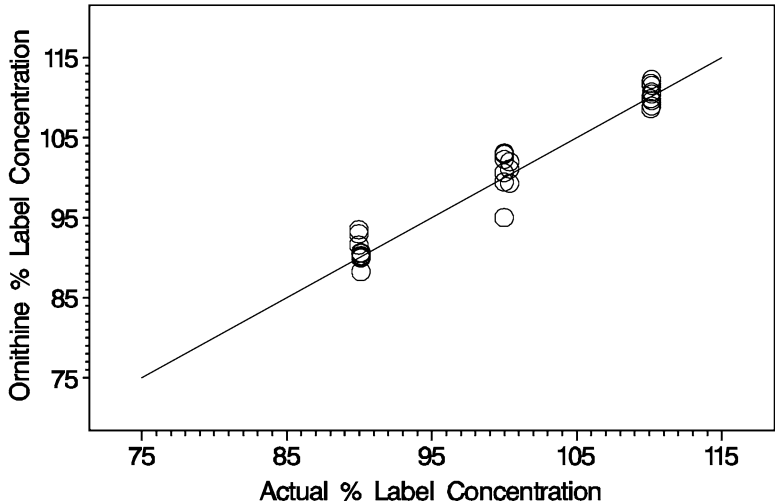


Fig. 2.5 Observed measures versus target values of ornithine

Figures 2.4–2.7 present the plots and Tables 2.4–2.7 present the agreement statistics for glycine, ornithine, L-threonine acid, and L-methionine, respectively.

The results for glycine are accurate and precise, with $CCC = 0.996$ (0.994), $r = 0.998$ (0.996), $c_a = 0.998$ (0.996), $TDI_{0.8} = 0.93$ (1.18), and $CP_3 = 0.9999$ (0.9966). Values presented in parentheses represent the respective 95% lower or upper confidence limit. More than 80% of observations are within 0.93 of target values. The 95% upper confidence limit of $TDI_{0.8}$ is 1.18, which is within the

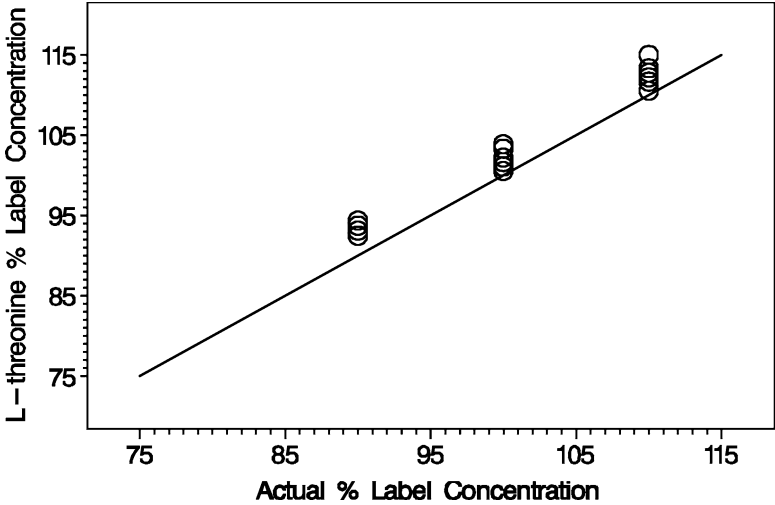


Fig. 2.6 Observed measures versus target values of L-threonine

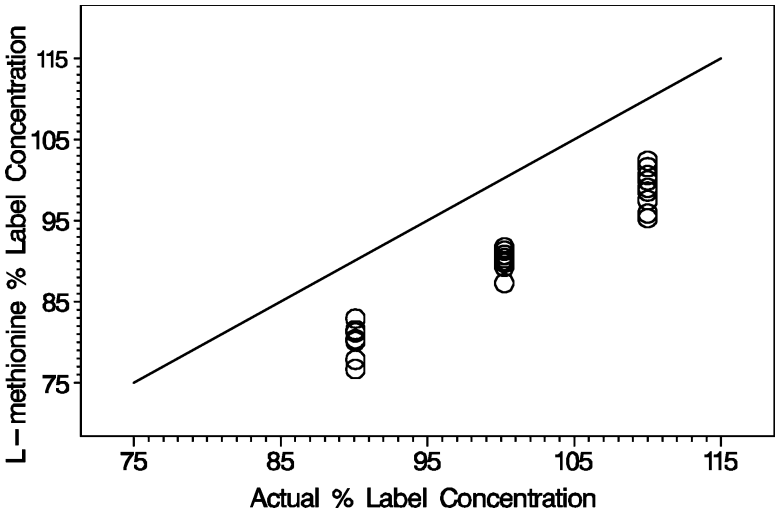


Fig. 2.7 Observed measures versus target values of L-methionine

Table 2.4 Glycine results ($n = 27$)

Statistics	CCC	Precision coefficient	Accuracy coefficient	$TDI_{0.8}$	CP_3	RBS
Estimate	0.9962	0.9981	0.9981	0.93	0.9999	0.08
95% Conf. limit	0.9937	0.9963	0.9961	1.18	0.9966	–
Allowance	0.9900	–	–	3.0	0.8000	–

“–” means “not applicable”

Table 2.5 Ornithine results ($n = 27$)

Statistics	CCC	Precision coefficient	Accuracy coefficient	$TDI_{0.8}$	CP_3	RBS
Estimate	0.9738	0.9759	0.9978	2.45	0.8703	0.08
95% Conf. limit	0.9502	0.9535	0.9801	3.09	0.7496	–
Allowance	0.9900	–	–	3.0	0.8000	–

“–” means “not applicable”

Table 2.6 L-threonine results ($n = 27$)

Statistics	CCC	Precision coefficient	Accuracy coefficient	$TDI_{0.8}$	CP_3	RBS
Estimate	0.9444	0.9905	0.9534	3.61	0.6557	4.44
95% Conf. limit	0.9084	0.9814	0.9222	4.14	0.5188	–
Allowance	0.9900	–	–	3.0	0.8000	–

“–” means “not applicable”

Table 2.7 L-methionine results ($n = 27$)

Statistics	CCC	Precision coefficient	Accuracy coefficient	$TDI_{0.8}$	CP_3	RBS
Estimate	0.5308	0.9723	0.5459	13.68	0.0001	26.11
95% Conf. limit	0.3991	0.9464	0.4280	14.89	0.0000	–
Allowance	0.9900	–	–	3.0	0.8000	–

“–” means “not applicable”

allowable 3%. The 95% lower confidence limit of CP_3 is 0.997, which is better than the allowable 0.8. The CCC estimate is near 1, indicating an almost perfect agreement.

The results of ornithine are accurate but less precise, with $CCC = 0.974$ (0.95), $r = 0.976$ (0.954), $c_a = 0.998$ (0.98), $TDI_{0.8} = 2.45$ (3.09), and $CP_3 = 0.870$ (0.750). The 95% upper confidence limit of $TDI_{0.8}$ is 3.09, and the 95% lower confidence limit of CP_3 is 0.750.

The results of L-threonine are inaccurate but precise, with $CCC = 0.944$ (0.908), $r = 0.991$ (0.981), $c_a = 0.953$ (0.922), $TDI_{0.8} = 3.61$ (4.14), and $CP_3 = 0.656$ (0.519). The 95% upper confidence limit of $TDI_{0.8}$ is 4.14, and the 95% lower confidence limit of CP_3 is 0.519.

The results of L-methionine are inaccurate and imprecise, with $CCC = 0.531$ (0.399), $r = 0.972$ (0.946), $c_a = 0.546$ (0.428), $TDI_{0.8} = 13.68$ (14.89), and $CP_3 = 0.0001$ (0.0000). The 95% upper confidence limit of $TDI_{0.8}$ is 14.89, and the 95% lower confidence limit of CP_3 is almost zero. Note that the $TDI_{0.8}$ estimate of the L-methionine assay is conservative, since the estimate of its relative bias squared value is large.

In summary, only the glycine assay in this example meets the client’s criterion.

2.8.4 Example 4: Lab Performance Process Control

This example was presented in Lin (2008). For quality control of clinical laboratories, control materials of various concentrations were randomly sent to laboratories for testing. The test results were to satisfy the proficient testing (PT) criterion. The PT criterion for each lab test of the Clinical Laboratory Improvement Amendments (CLIA) Final Rule (2003) <http://wwwn.cdc.gov/clia/regs/subpart.i.aspx#493.929> required that 80% of observations be within a certain percentage or unit of the target concentration for measuring control materials. The target concentrations usually were the average of control materials across a peer group of labs using similar instruments. Such a criterion lends itself directly for using the $TDI_{0.8}$ or $TDI_{0.8}$.

For each of the majority of lab measurements, laboratories were required to test commercial control materials at least once a day for at least two concentrations (low and high). Daily glucose values of 116 laboratory instruments were monitored. Based on accuracy and precision indices, we selected four laboratory instruments with four distinct combinations of precision and accuracy. For each laboratory instrument we computed $TDI_{0.9}$, CCC, and precision and accuracy coefficients. Here, we chose to monitor 90% for a cushion, instead of 80%, of observations across all levels that were within $TDI_{0.9}$ or $TDI_{0.9}$ units of targets. We can translate from $TDI_{0.9}$ to $TDI_{0.8}$ by multiplying by $1.282/1.645 = 0.779$. The target values were computed as the average of these 116 laboratory instruments. For glucose, this PT criterion was 10% or 6 mg/dL, whichever was larger. The range of these data was around 70–270 mg/dL. In this case, the 10% value was the PT criterion (always larger than 6 mg/dL).

For each lab instrument, we computed the preceding agreement statistics for each calendar month and for the last available 50 days (current window). Across the 116 lab instruments, we computed the group geometric mean (GM), one standard deviation (1-SD), and two standard deviation (2-SD) upper limits with 3-month average $TDI_{0.9}$ values as benchmarks. Note that the distribution of $TDI_{0.9}$ was shown to be log-normal. Here, the confidence limit of $TDI_{0.9}$ per lab instrument was not computed, because we were using the population benchmarks. Therefore, it is irrelevant whether the target values were treated as random or fixed.

Figures 2.8–2.11 present the plots of the four selected cases. For each case, the left-hand plot presents the usual agreement plot of observations versus target values for the current window, and each plotted symbol (circle) represents the daily glucose value against the daily average glucose values across 116 labs. The right-hand plot monitors the quality control results over a selected time window based on $TDI_{0.9}$ values. We chose to monitor a rolling three-completed-month window (June, July, and August in this case) plus the current window. Each plotted symbol (dot) represents the monthly or current window $TDI_{0.9}$ value. Also presented are the population benchmarks of geometric mean, 1-SD, and 2-SD upper limits, and the PT criterion (PTC) of 10% (dashed line). Although the CCC, precision coefficient, and accuracy coefficient are not shown in the right-hand plot, these

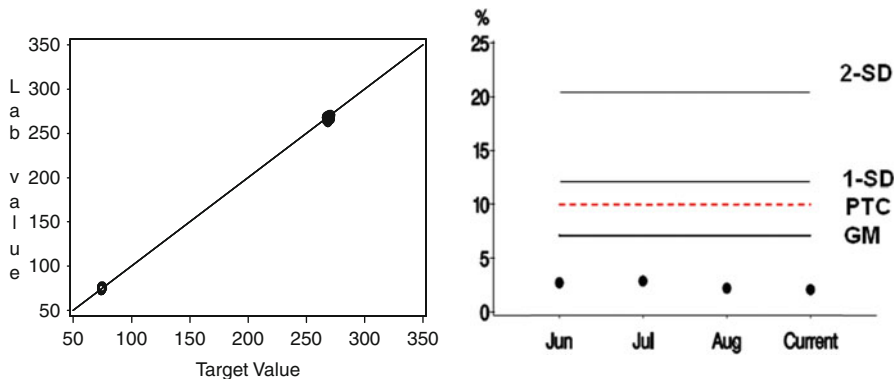


Fig. 2.8 Observed glucose measures versus target values of lab instrument for the current window and the control chart based on $TDI\%_{0.9}$: almost perfect

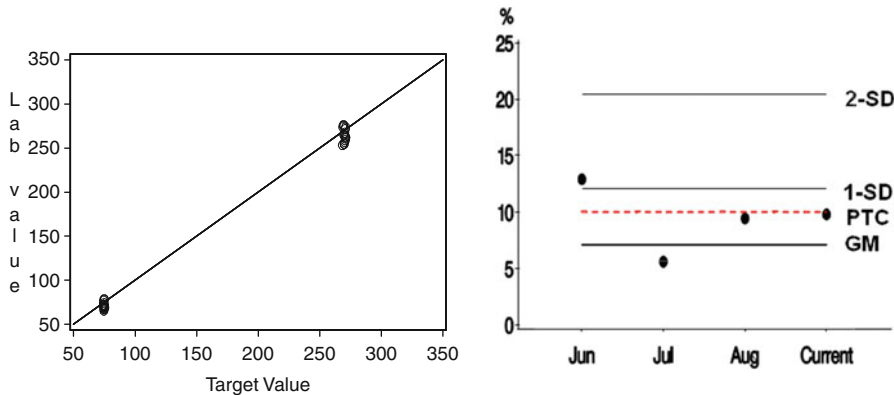


Fig. 2.9 Observed glucose measures versus target values of lab instrument for the current window and the control chart based on $TDI\%_{0.9}$: imprecise but accurate

values of the current window were used to select the four instruments that are presented here. The use of CP can be helpful here. However, the CP values have difficulty discriminating among good instruments when they all have very high CP values.

Figure 2.8 shows the best-performing lab instrument among all 116 lab instruments, with $CCC = 0.9998$, $r^2 = 0.9997$, $c_a = 0.9999$, and $TDI\%_{0.9} = 2.1\%$ for the current window. It has an almost perfect CCC, and its $TDI\%_{0.9}$ values are around 2%–3%.

Figure 2.9 shows a less-precise but accurate lab instrument, with $CCC = 0.996$, $r^2 = 0.996$, $c_a = 0.998$, and $TDI\%_{0.9} = 9.8\%$ for the current window. Its values rank at around 2/3 (slightly greater than 1-SD value) among its peers in June, slightly better than its peer average in July, and at around the PTC level in August and the current window.

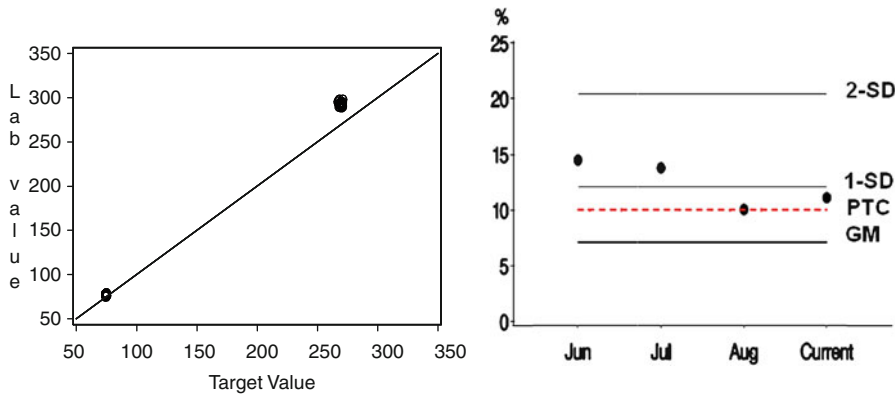


Fig. 2.10 Observed glucose measures versus target values of lab instrument for the current window and the control chart based on $TDI\%_{0.9}$: precise but inaccurate

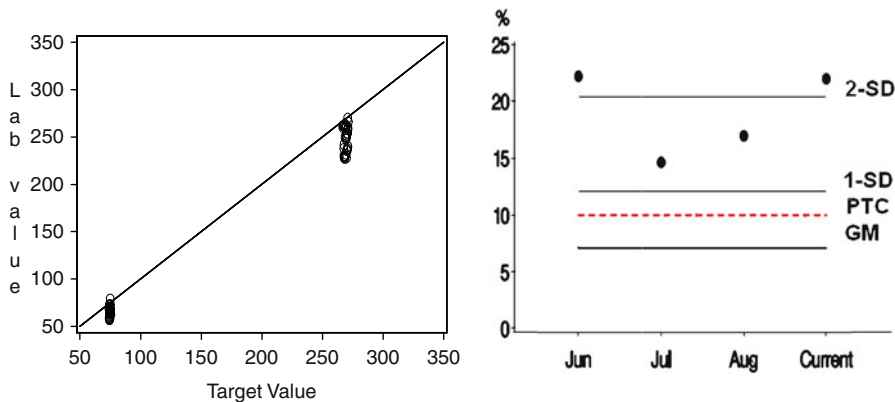


Fig. 2.11 Observed glucose measures versus target values of lab-instrument for the current window and the control chart based on $TDI\%_{0.9}$: imprecise and inaccurate

Figure 2.10 shows a precise but inaccurate lab instrument, with $CCC = 0.995$, $r^2 = 0.9996$, $c_a = 0.995$, and $TDI\%_{0.9} = 11\%$ for the current window. Its $TDI\%_{0.9}$ values are between 1-SD and 2-SD values of its peers in June and July, at around the PTC level in August, and slightly worse than the PTC in the current window.

Figure 2.11 shows the worst-performing lab instrument among all 116 lab instruments, with $CCC = 0.983$, $r^2 = 0.991$, $c_a = 0.988$, and $TDI\%_{0.9} = 22\%$ for the current window. Its $TDI\%_{0.9}$ values are between 1-SD and 2-SD values of its peers in July and August, and worse than the 2-SD value of its peers in June and the current window.

This example conveys a few lessons. First, it is difficult to judge agreement solely by the CCC, precision coefficient, and accuracy coefficient in their absolute values. All comparisons should be judged in relative terms. Here, even the worst CCC value shown in Fig. 2.11 is 0.983. Such a high value here is due to the large study range of

70–270 mg/dL. Note that as stated earlier, the CCC, precision coefficient, accuracy coefficient, and ICC depend largely on the study range. Comparisons among any of these are valid only with similar study ranges. In this example, the study ranges (based on peer means) are identical. It is important to report the study range when reporting these statistics.

Second, not all lab instruments are created equal. Their quality, in terms of $\text{TDI}_{0.9}$ values, could range from 2% to 22%, which is quite diverse. It is important to submit our blood samples to a lab with a good reputation for quality.

Third, the PTC value is between the GM and 1-SD benchmarks for measuring glucose, which means that about one-third of the lab instruments are in danger of failing the PTC as dictated by the CLIA 2003 Final Rule. Perhaps the PTC is set too strictly for measuring glucose. Note that we use $\text{TDI}_{0.9}$ instead of $\text{TDI}_{0.8}$ values for a cushion here.

2.8.5 Example 5: Clinical Chemistry and Hematology Measurements That Conform to CLIA Criteria

The data for this example were obtained through the clinical labs within the research and development organization at Baxter Healthcare Corporation. Analysis of serum chemistry analytes previously validated using the Hitachi 911 (Roche Diagnostics) chemistry analyzer (reference or gold standard assays) was to be converted to the Olympus AU400e (Olympus America, Inc.) chemistry analyzer (test or new assays). Assay comparisons were performed by analyzing approximately 50–60 samples per assay on both analyzers.

Hematology analyses were performed using two automated hematology instruments: the CELL DYN 3500 (Abbott Diagnostics) as the reference and the ADVIA 2120 (SIEMENS Healthcare Diagnostics) as the test. Analyses were performed on whole blood samples drawn into tubes containing EDTA anticoagulant. A total of 93 samples from 16 humans, 18 rabbits, 19 rats, 20 dogs, and 20 pigs were each tested once on both instruments. All species were combined to establish wider data ranges, and assay performance was not expected to depend on species.

Evaluations included clinical chemistry analytes of albumin, BUN, calcium, chloride, creatinine, glucose, iron, magnesium, potassium, sodium, total protein, and triglycerides; and hematology analytes of hemoglobin, platelet count, RBC, and WBC. Any analyte without a PTC or values outside the data range was not included in the evaluations.

Table 2.8 presents the data range, PT criterion, and estimates and confidence limits (in parentheses) of agreement statistics for each of the above analytes. Data ranges of clinical chemistry analytes were acquired from the Olympus Chemistry Reagent Guide “Limitations of the Procedure” section for each assay (Olympus America, Inc. Reagent Guide Version 3.0. May 2007. Irving, TX). Data ranges of hematology analytes were acquired from ADVIA 2120 Version 5.3.1–MS2005 Bayer (now Siemens) Healthcare LLC.

Table 2.8 Agreement statistics against the PT criteria (PTC) for clinical chemistry and hematology analytes

Analyte	Range	PTC	n	CCC	Precision Coef.	Accuracy Coef.	TDI _{0.8} ^a	CP _{PTC} ^b	RBS ^c
Albumin (g/dL)	1.5–6.0	10%	58	0.723 (0.626)	0.859 (0.789)	0.842 (0.768)	23.6% (27.29%)	0.379 (0.309)	1.19
BUN ≤ 22 (mg/dL)	2–22	2	44	0.993 (0.989)	0.996 (0.993)	0.997 (0.994)	0.77 (0.91)	1.000 (0.996)	0.53
BUN > 22 (mg/dL)	22–130	9%	16	0.999 (0.998)	0.999 (0.998)	1.000 (0.999)	3.26% (4.48%)	0.999 (0.954)	0.05
Calcium (mg/dL)	4–18	1	60	0.873 (0.833)	0.996 (0.994)	0.876 (0.838)	1.58 (1.69)	0.302 (0.223)	11.7
Chloride (mmol/L)	50–200	5%	61	0.991 (0.987)	0.996 (0.993)	0.995 (0.992)	2.44% (2.8%)	0.995 (0.982)	0.81
Creatinine Enz < 2 (mg/dL)	0.2–2	0.3	50	0.989 (0.982)	0.990 (0.983)	0.999 (0.996)	0.08 (0.09)	1.000 (1.000)	0.06
Creatinine Jaffe < 2 (mg/dL)	0.2–2	0.3	47	0.963 (0.942)	0.981 (0.969)	0.981 (0.966)	0.14 (0.16)	0.998 (0.989)	0.92
Glucose > 60 (mg/dL)	60–800	10%	55	0.999 (0.998)	0.999 (0.998)	1.000 (0.999)	4.28% (5.01%)	0.997 (0.986)	0.32
Iron (ug/dL)	10–1000	20%	59	0.994 (0.991)	0.997 (0.995)	0.997 (0.995)	9.74% (11.45%)	0.987 (0.961)	0.03
Magnesium (mg/dL)	0.5–8	25%	61	0.970 (0.958)	0.996 (0.993)	0.974 (0.963)	14.41% (15.71%)	0.999 (0.995)	5.82
Potassium (mmol/L)	1–10	0.5	59	0.996 (0.995)	0.998 (0.997)	0.998 (0.997)	0.14 (0.16)	1.000 (1.000)	0.43
Sodium (mmol/L)	50–200	4	59	0.994 (0.991)	0.995 (0.993)	0.999 (0.997)	1.77 (2.06)	0.996 (0.984)	0.16
Total Protein (g/dL)	3–12	10%	56	0.993 (0.989)	0.993 (0.989)	1.000 (0.997)	2.31% (2.71%)	1.000 (1.000)	0.03
Triglycerides (g/dL)	10–1000	25%	56	0.997 (0.996)	0.998 (0.998)	0.999 (0.998)	8.2% (9.56%)	1.000 (0.999)	0.59
Hemoglobin (g/dL)	1–22.5	7%	93	0.967 (0.957)	0.997 (0.996)	0.969 (0.961)	6.02% (6.4%)	0.942 (0.903)	7.08
Platelet Count (×103/ L)	10–3500	25%	84	0.917 (0.884)	0.926 (0.896)	0.990 (0.973)	31.72% (36.79%)	0.695 (0.629)	0.04
RBC (×106/ μL)	0.1–12	6%	92	0.988 (0.984)	0.997 (0.995)	0.992 (0.989)	4.13% (4.52%)	0.966 (0.937)	2.22
WBC (×103/ L)	0.1–100	15%	93	0.942 (0.922)	0.958 (0.941)	0.984 (0.972)	21.34% (24.41%)	0.640 (0.576)	0.08

Note: Shown in parentheses is the 95% upper (TDI) or lower (CCC, precision, accuracy, CP) confidence limit. Boldface analytes are those that failed the PTC.

^a Total deviation index to cover 80% of absolute difference or % change. The 95% upper limit should be less than PTC or PTC%.

^b Coverage probability of values within the PTC. The 95% lower limit should be greater than 0.8.

^c The relative bias squared (RBS) must be less than 8 in order for the approximate TDI to be valid. Otherwise, the TDI estimate is conservative depending on the RBS value.

For evaluation of BUN, the PTC is 9% for values greater than or equal to 22 mg/dL or 2 mg/dL for values less than or equal to 22 mg/dL, since 9% of 22 is approximately 2. The criterion used for creatinine Enz and Jaffe is 0.3 mg/dL, since only values less than 2 mg/dL were evaluated. Glucose was evaluated for values greater than 60 mg/dL with the criterion 10%.

Figure 2.12 presents the agreement plots of the above analytes. The analytes of albumin, calcium, platelet count, and WBC had precision and/or accuracy problems, while the other analytes appeared to perform well. Comparing the 95% upper confidence limit of TDI against the PTC or PTC% or the 95% lower confidence limit CP against 0.8, all but the analytes of albumin, calcium, platelet count, and WBC (shown boldface in Table 2.8) pass the PTC with 95% confidence.

Table 2.9 presents the results of traditional statistical analyses based on paired *t*-test and ordinary regression. The results from Deming (orthogonal) regression by treating *X* as a random variable are not shown because they are more or less similar to those of ordinary regression. Table 2.9 shows the data range, sample size, paired *t*-test *p*-value, intercepts, slopes, and the testing of intercept (0) and slope (1) of ordinary regressions.

The paired *t*-test rejects the agreement of extremely well performing analytes, that is, $\text{BUN} \leq 22$ mg/dL, chloride, creatinine Jaffe, glucose, magnesium, potassium, sodium, triglycerides, hemoglobin, and RBC, with $p = 0.003$ for sodium and $p < 0.001$ for the others. These rejections correspond to the left-hand plot of Fig. 1.2 and are due primarily to near-zero residual variance and/or large sample size. On the other hand, the paired *t*-test accepts ($p = 0.062$) the agreement of platelet count. Such failure to reject corresponds to the right-hand plot of Fig. 1.2, due primarily to its large residual variance.

In terms of ordinary regression, tests of intercept (0) and/or slope (1) (LS01) reject ($p < 0.05$) the agreement of the extremely well performing analytes of $\text{BUN} \leq 22$ mg/dL, chloride, creatinine Jaffe, iron, magnesium, potassium, sodium, triglycerides, and hemoglobin. These rejections are similar, although not identical, to the paired *t*-test for the same reasons. On the other hand, LS01 accepts ($p \geq 0.373$) the agreement of platelet count for the same reasons as paired *t*-test.

In terms of CCC, precision and accuracy for clinical chemistry analysts of BUN, chloride, glucose, iron, magnesium, potassium, sodium, total protein, triglycerides have excellent agreement ($\text{CCC} > 0.99$) between measurements from the Olympus AU400e and Hitachi 911 instruments. Both have excellent precision (>0.99) and accuracy (>0.99).

Creatinine Enz and creatinine Jaffe comfortably pass the PTC. Because their data ranges are small, from 0.2 to 1.8 mg/dL, their CCC and precision and accuracy coefficients become relatively lower. Magnesium also pass the PTC by a comfortable margin. It has excellent precision (0.9956) but relatively smaller accuracy (0.9738), because most Olympus values are smaller than Hitachi values by a negligible amount.

For clinical chemistry analytes of albumin and calcium, the lab has difficulties proving the equivalence of measurements from the Olympus AU400e and Hitachi 911 instruments. Albumin measurements from both the Olympus AU400e and Hitachi 911 are neither accurate (0.8422) nor precise (0.8589). We are 95%

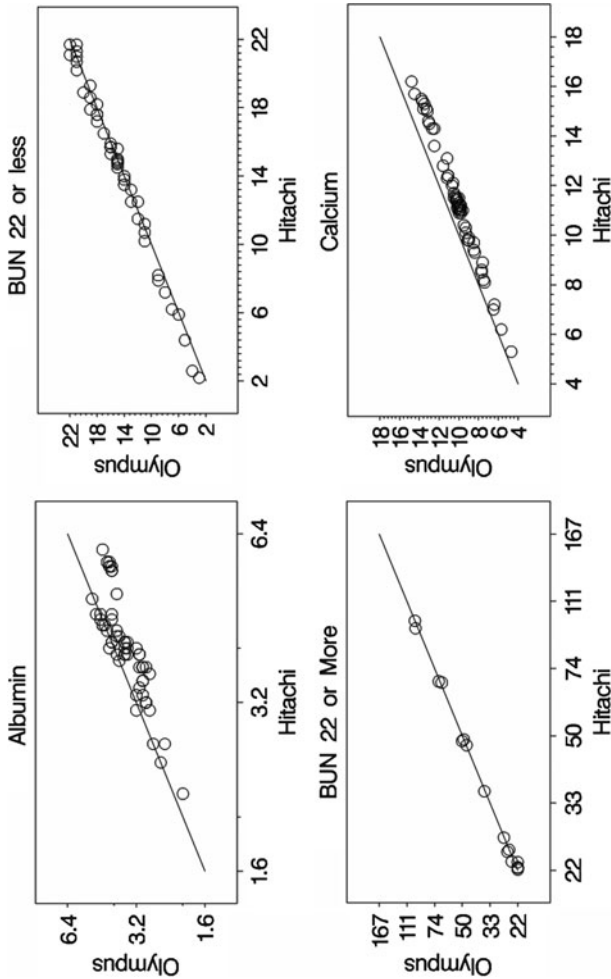


Fig. 2.12 Agreement plots

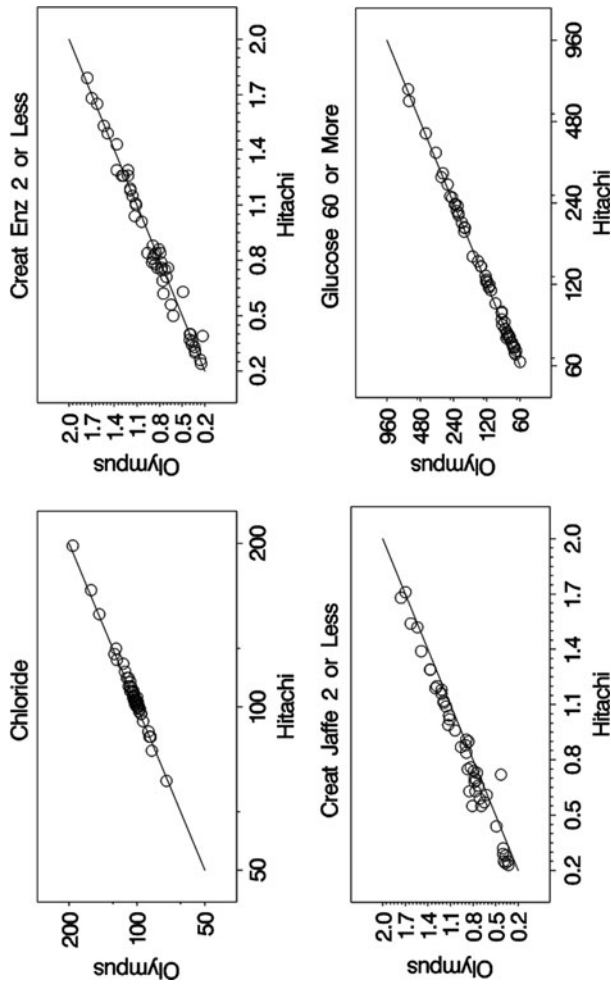


Fig. 2.12 (continued)

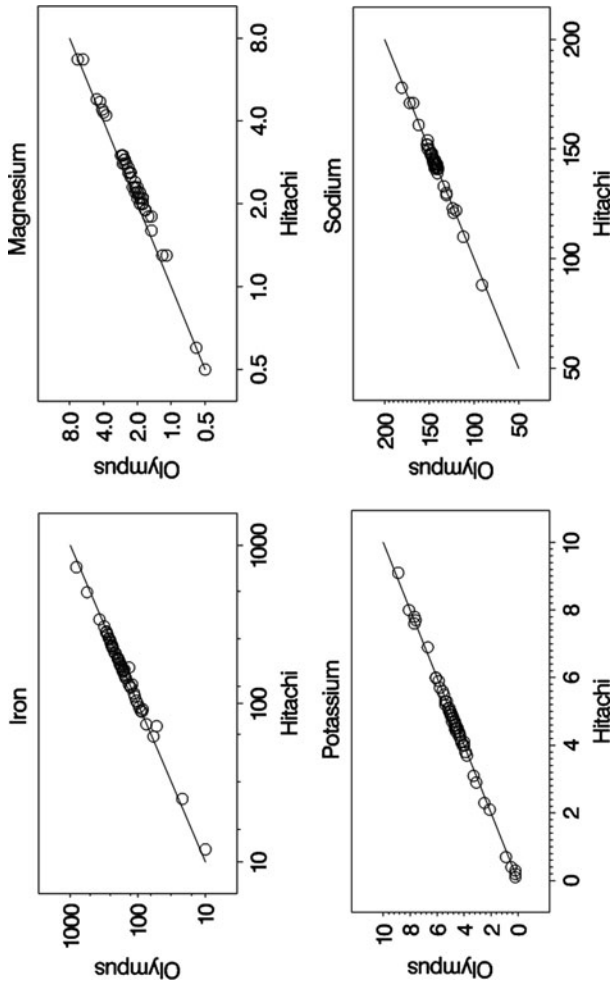


Fig. 2.12 (continued)

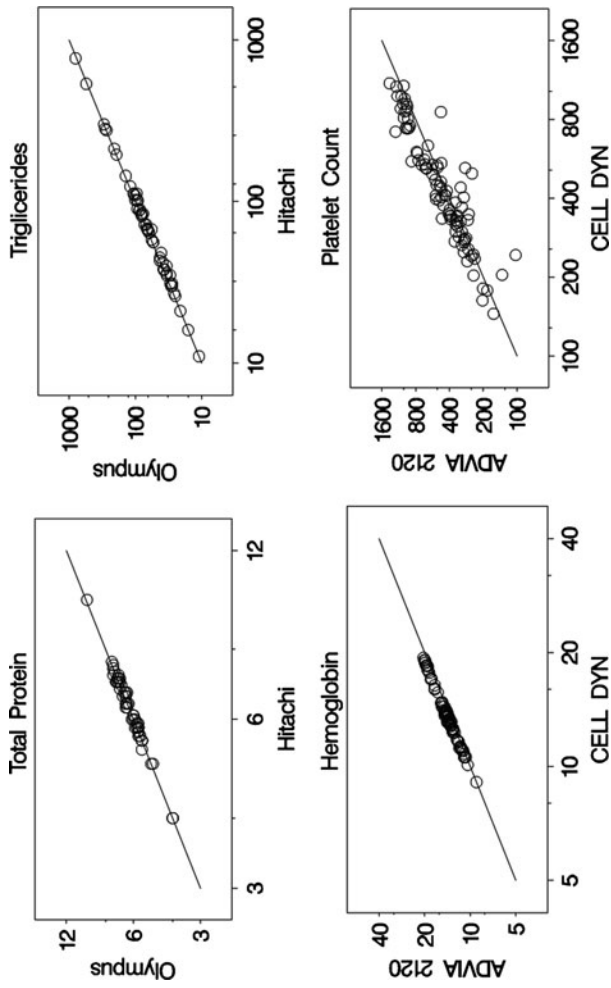


Fig. 2.12 (continued)

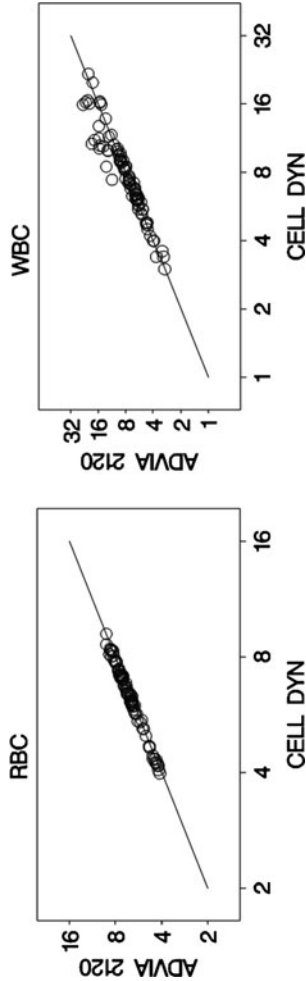


Fig. 2.12 (continued)

Table 2.9 Traditional statistics for clinical chemistry and hematology analytes

Analyte	Range	n	Paired <i>t</i> -test		Intercept	Slope	Intercept = 0		Slope = 1	
			<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value
Albumin (g/dL)	1.5–6.0	58	<0.001		0.181	0.783	0.045		0.001	
BUN ≤ 22 (mg/dL)	2–22	44	<0.001		0.848	0.966	0		0.014	
BUN > 22 (mg/dL)	22–130	16	0.365		–0.037	1.012	0.401		0.329	
Calcium (mg/dL)	4–18	60	<0.001		0.15	0.884	0.205		<0.001	
Chloride (mmol/L)	50–200	61	<0.001		0.16	0.963	0.005		0.002	
Creatinine Enz < 2 (mg/dL)	0.2–2	50	0.093		0.026	0.986	0.196		0.517	
Creatinine Jaffe < 2 (mg/dL)	0.2–2	47	<0.001		0.107	0.962	0		0.186	
Glucose > 60 (mg/dL)	60–800	55	<0.001		–0.049	1.007	0.129		0.306	
Iron (ug/dL)	10–1000	59	0.214		–0.383	1.078	0		<0.001	
Magnesium (mg/dL)	0.5–8	61	<0.001		–0.095	0.999	0		0.907	
Potassium (mmol/L)	1–10	59	<0.001		0.22	0.967	0		<0.001	
Sodium (mmol/L)	50–200	59	0.003		4.025	0.975	0.032		0.06	
Total Protein (g/dL)	3–12	56	0.198		0.015	0.99	0.602		0.531	
Triglycerides (g/dL)	10–1000	56	<0.001		–0.104	1.016	0.003		0.049	
Hemoglobin (g/dL)	1–22.5	93	<0.001		–0.078	1.046	0.001		<0.001	
Platelet Count (×103/μL)	10–3500	84	0.062		–0.211	1.042	0.461		0.373	
RBC (×106/μL)	0.1–12	92	<0.001		0.005	1.012	0.776		0.194	
WBC (×103/μL)	0.1–100	93	0.006		–0.196	1.116	0.009		0.001	

confident that albumin measurements can deviate 27.3% ($>10\%$) from their target values, and that the measured values conformed to the PTC only 30.9% ($<80\%$) of the time. Calcium measurements are precise (0.9963) but not accurate (0.8759). We are 95% confident that calcium measurements can deviate 1.58 mg/dL (>1 mg/dL) from their target values, and that the measured values conform to the PTC only 22.8% ($<80\%$) of the time.

For hemoglobin, the lab has good agreement ($CCC = 0.9665$) between measurements from the Advia 2120 and Cell DYN 3500 instruments, with excellent precision (0.9970) and good accuracy (0.9694). There is a small bias showing that measurements from the Advia 2120, with only one exception, are consistently higher than from the Cell DYN 3500. Note that data were not collected over the full analytical range of 5 to 22 g/dL for this analyte.

RBC counts have very good agreement ($CCC = 0.9883$) between measurements from the Advia 2120 and Cell DYN 3500 instruments. Both have excellent precision (0.9965) and accuracy (0.9917). There is a small bias showing that all but one of the measurements from Advia 2120 are consistently higher than those of Cell DYN 3500. Note that data are not collected over the full analytical range of 2 to $10 \times 10^6/\mu\text{L}$ for this analyte.

Platelet count from the Advia 2120 and Cell DYN 3500 are accurate (0.9897) but imprecise (0.9263). We are 95% confident that platelet count measurements can deviate 36.8% ($>25\%$) from their target values and that they conformed to the PTC only 62.9% ($<80\%$) of the time. Additionally, WBC measurements are also relatively accurate (0.9839) but imprecise (0.9576), especially for readings greater than $8 \times 10^3/\mu\text{L}$. The lab fails to show that these two analytes meet the PTC. We are 95% confident that WBC measurements can deviate 24.4% ($>15\%$) from their target values, and that they conform to the PTC only 58.0% ($<80\%$) of the time.

In summary, using the agreement statistics presented in this example, 14 out of 18 method comparison cases meet the CLIA criteria with 95% confidence. Of the four that did not meet the CLIA criteria, one is acceptable by the traditional paired t -test and regression analysis. Of the 14 that meet the CLIA criteria, 11 are rejected by the traditional paired t -test or regression analysis.

2.9 Proofs of Asymptotical Normality When Target Values Are Random

2.9.1 CCC and Precision Estimates

This proof can be seen in Lin (1989). The Z transformation of the CCC estimate can be expressed as $Z = g(\mathbf{m})$, where

$$\begin{aligned} \mathbf{m} &= (m_1, m_2, m_3, m_4, m_5)' \\ &= \left(\bar{Y}, \bar{X}, \frac{1}{n} \sum_{i=1}^n X_i^2, \frac{1}{n} \sum_{i=1}^n Y_i^2, \frac{1}{n} \sum_{i=1}^n Y_i X_i \right)', \end{aligned} \quad (2.43)$$

and

$$Z = g(\mathbf{m}) = \frac{1}{2} \ln \left[1 + \frac{4(m_5 - m_1 m_2)}{m_3 + m_4 - 2m_5} \right].$$

The vector \mathbf{m} is expressed as a function of sample moments, and has an asymptotic 5-variate normality with mean

$$\Theta = (\mu_y, \mu_x, \sigma_y^2 + \mu_y^2, \sigma_x^2 + \mu_x^2, \sigma_{yx} + \mu_y \mu_x)',$$

and variance $\frac{1}{n} \Sigma$, where

$$\Sigma = \{\lambda_{ij}\}_{5 \times 5}. \quad (2.44)$$

Here, with the assumption that $\{(Y_i, X_i) \mid i = 1, 2, \dots, n\}$ are random samples from a bivariate normal distribution, we have

$$\begin{aligned} \lambda_{11} &= \sigma_y^2, \\ \lambda_{12} &= \lambda_{21} = \sigma_{yx}, \\ \lambda_{22} &= \sigma_x^2, \\ \lambda_{13} &= \lambda_{31} = 2\mu_y \sigma_y^2, \\ \lambda_{23} &= \lambda_{32} = 2\mu_y \sigma_{yx}, \\ \lambda_{33} &= 2\sigma_y^4 + 4\sigma_y^2 \mu_y^2, \\ \lambda_{14} &= \lambda_{41} = 2\mu_x \sigma_{yx}, \\ \lambda_{24} &= \lambda_{42} = 2\mu_x \sigma_x^2, \\ \lambda_{34} &= \lambda_{43} = 2\sigma_{yx}^2 + 4\mu_y \mu_x \sigma_{yx}, \\ \lambda_{44} &= 2\sigma_x^4 + 4\sigma_x^2 \mu_x^2, \\ \lambda_{15} &= \lambda_{51} = \mu_x \sigma_y^2 + \mu_y \sigma_{yx}, \\ \lambda_{25} &= \lambda_{52} = \mu_y \sigma_x^2 + \mu_x \sigma_{yx}, \\ \lambda_{35} &= \lambda_{53} = 2\sigma_{yx} \mu_y^2 + 2\sigma_{yx} \sigma_y^2 + 2\mu_y \mu_x \sigma_y^2, \\ \lambda_{45} &= \lambda_{54} = 2\sigma_{yx} \mu_x^2 + 2\sigma_{yx} \sigma_x^2 + 2\mu_y \mu_x \sigma_x^2, \end{aligned}$$

and

$$\lambda_{55} = \sigma_y^2 \sigma_x^2 + \mu_y^2 \sigma_x^2 + \mu_x^2 \sigma_y^2 + \sigma_{yx}^2 + 2\mu_y \mu_x \sigma_{yx}.$$

It follows from the delta method or from the theory of functions of asymptotically normal vectors (Serfling 1980, Corollary 3.3) that Z is asymptotically normal with mean $\frac{1}{2} \ln \frac{1+\rho_c}{1-\rho_c}$ and variance $\frac{1}{n} \mathbf{d}' \Sigma \mathbf{d}$, where

$$\begin{aligned} \mathbf{d} &= (d_1, d_2, d_3, d_4, d_5)' \\ &= \left(\left. \frac{\partial g(\mathbf{m})}{\partial m_1} \right|_{\mathbf{m}=\Theta}, \dots, \left. \frac{\partial g(\mathbf{m})}{\partial m_5} \right|_{\mathbf{m}=\Theta} \right)'. \end{aligned}$$

The elements of d are

$$\begin{aligned} d_1 &= \frac{-\mu_x}{\sigma_y^2 + \sigma_x^2 + 2\sigma_{yx} + (\mu_y - \mu_x)^2}, \\ d_2 &= \frac{-\mu_y}{\sigma_y^2 + \sigma_x^2 + 2\sigma_{yx} + (\mu_y - \mu_x)^2}, \\ d_3 &= d_4 = \frac{-\sigma_{yx}}{\left[\sigma_y^2 + \sigma_x^2 + 2\sigma_{yx} + (\mu_y - \mu_x)^2 \right] \left[\sigma_y^2 + \sigma_x^2 - 2\sigma_{yx} + (\mu_y - \mu_x)^2 \right]}, \end{aligned}$$

and

$$d_5 = \frac{(\sigma_y^2 + \sigma_x^2) + (\mu_y - \mu_x)^2}{\left[\sigma_y^2 + \sigma_x^2 + 2\sigma_{yx} + (\mu_y - \mu_x)^2 \right] \left[\sigma_y^2 + \sigma_x^2 - 2\sigma_{yx} + (\mu_y - \mu_x)^2 \right]}.$$

After straightforward, albeit tedious algebraic calculations, it can be shown that the variance of Z is

$$\begin{aligned} \sigma_Z^2 &= \frac{1}{n} \mathbf{d}' \mathbf{\Sigma} \mathbf{d} \\ &= \frac{1}{n} \left[\frac{(1 - \rho^2) \rho_c^2}{(1 - \rho_c^2) \rho^2} + \frac{2\rho_c^3(1 - \rho_c)v^2}{\rho(1 - \rho_c^2)^2} - \frac{\rho_c^4 v^4}{2\rho^2(1 - \rho_c^2)^2} \right]. \end{aligned} \quad (2.45)$$

The Z -transformed CCC estimate can approach normality much more rapidly as confirmed by the Monte Carlo experiment in Lin (1989). When $\rho_c = \rho$ and $v = 0$, (2.45) degenerates into $\frac{1}{n}$, which is the variance of the Z transformation of the precision estimate.

2.9.2 MSD Estimate

From (2.43), we can write the natural log transformation of the MSD estimate, or $W = \ln(e^2)$ as

$$W = g(m_3, m_4, m_5) = \ln(m_3 + m_4 - 2m_5).$$

By the delta method, W is asymptotically normal with mean $\ln(\epsilon^2)$ and variance $\frac{1}{n} \mathbf{d}' \mathbf{\Sigma} \mathbf{d}$, where

$$\mathbf{\Sigma} = \begin{pmatrix} \lambda_{33} & \lambda_{34} & \lambda_{35} \\ & \lambda_{44} & \lambda_{45} \\ & & \lambda_{55} \end{pmatrix}$$

and

$$\underline{d} = \left(\frac{\partial g(\mathbf{m})}{\partial m_3} \Big|_{\mathbf{m}=\Theta}, \frac{\partial g(\mathbf{m})}{\partial m_4} \Big|_{\mathbf{m}=\Theta}, \frac{\partial g(\mathbf{m})}{\partial m_5} \Big|_{\mathbf{m}=\Theta} \right)' = \left(\frac{1}{\varepsilon^2}, \frac{1}{\varepsilon^2}, -\frac{2}{\varepsilon^2} \right)'.$$

After straightforward algebraic calculations, we can show that the variance of W is

$$\sigma_W^2 = \frac{2}{n} \left[1 - \frac{(\mu_y - \mu_x)^4}{\varepsilon^4} \right]. \quad (2.46)$$

2.9.3 CP Estimate

This proof can be seen in Lin, Hedayat, Sinha, and Yang (2002). Here, we use a different approach to demonstrating the delta method. We can use a first-order approximation to compute the mean and variance of the CP estimate p_{δ_0} by

$$\begin{aligned} p_{\delta_0} &= \Phi \left(\frac{\delta_0 - \mu_d}{\sigma_d} \right) - \frac{1}{\sigma_d} \phi \left(\frac{\delta_0 - \mu_d}{\sigma_d} \right) (\bar{X}_d - \mu_d) \\ &\quad - \frac{\delta_0 - \mu_d}{\sigma_d^2} \phi \left(\frac{\delta_0 - \mu_d}{\sigma_d} \right) (s_d - \sigma_d) - \Phi \left(\frac{-\delta_0 - \mu_d}{\sigma_d} \right) \\ &\quad + \frac{1}{\sigma_d} \phi \left(\frac{-\delta_0 - \mu_d}{\sigma_d} \right) (\hat{\mu}_d - \mu_d) - \frac{\delta_0 + \mu_d}{\sigma_d^2} \phi \left(\frac{-\delta_0 - \mu_d}{\sigma_d} \right) (s_d - \sigma_d) \\ &\quad + O[(\bar{X}_d - \mu_d)^2] + O[(s_d - \sigma_d)^2] + O[(\bar{X}_d - \mu_d)(s_d - \sigma_d)], \quad (2.47) \end{aligned}$$

where $\phi(x)$ is the density function of the standard normal distribution, $\bar{X}_d = \bar{Y} - \bar{X}$, and $\lim_{x \rightarrow 0} \frac{O(x)}{x} < \infty$.

Therefore, the expected value of p_{δ_0} is

$$E(p_{\delta_0}) = \pi_{\delta_0} + O\left(\frac{1}{n}\right),$$

and the asymptotic variance of p_{δ_0} becomes

$$\begin{aligned} \sigma_{p_{\delta_0}}^2 &= \frac{1}{n} \left\{ \left[\phi \left(\frac{-\delta_0 - \mu_d}{\sigma_d} \right) - \phi \left(\frac{\delta_0 - \mu_d}{\sigma_d} \right) \right]^2 \right. \\ &\quad \left. + \frac{1}{2} \left[\frac{\delta_0 - \mu_d}{\sigma_d} \phi \left(\frac{\delta_0 - \mu_d}{\sigma_d} \right) + \frac{\delta_0 + \mu_d}{\sigma_d} \phi \left(\frac{-\delta_0 - \mu_d}{\sigma_d} \right) \right]^2 \right\} \\ &\quad + O\left(\frac{1}{n^2}\right). \quad (2.48) \end{aligned}$$

Because CP is bounded by 0 and 1, it is better to use the logit transformation for statistical inference. Let $T = \ln\left(\frac{p_{\delta_0}}{1-p_{\delta_0}}\right)$. Then the asymptotic mean of T is $\tau = \ln\left(\frac{\pi_{\delta_0}}{1-\pi_{\delta_0}}\right)$, and the asymptotic variance is $\sigma_T^2 = \frac{\sigma_{p_{\delta_0}}^2}{\pi_{\delta_0}^2(1-\pi_{\delta_0})^2}$.

2.9.4 Accuracy Estimate

This proof can be seen in Robieson (1999). This estimate, c_a , does not involve m_5 in (2.43). We redefine the new set of m vectors that are mostly uncorrelated as

$$\mathbf{m} = (m_1, m_2, m_3, m_4)' = (\bar{Y}, \bar{X}, s_y^2, s_x^2)',$$

which has an asymptotic 4-variate normality with mean $E(\mathbf{m}) = (\mu_y, \mu_x, \sigma_y^2, \sigma_x^2)'$ and variance $\frac{1}{n}\Sigma$, where $\Sigma = \{\lambda_{ij}\}_{4 \times 4}$.

Here,

$$\begin{aligned}\lambda_{11} &= \sigma_y^2, \\ \lambda_{12} &= \lambda_{21} = \sigma_{yx}, \\ \lambda_{13} &= \lambda_{14} = \lambda_{23} = \lambda_{24} = \lambda_{31} = \lambda_{41} = \lambda_{32} = \lambda_{42} = 0, \\ \lambda_{33} &= 2\sigma_y^4, \\ \lambda_{34} &= \lambda_{43} = 2\sigma_{yx}^2,\end{aligned}$$

and

$$\lambda_{44} = 2\sigma_x^4.$$

The logit transformation of the accuracy estimate can be written as

$$\begin{aligned}\mathbf{L} &= g(m_1, m_2, m_3, m_4) \\ &= \ln \left[\frac{2\sqrt{m_3 m_4}}{m_3 + m_4 + (m_1 - m_2)^2 - 2\sqrt{m_3 m_4}} \right],\end{aligned}$$

and is asymptotically normal with mean $\ln \frac{\chi_a}{1-\chi_a}$ and variance $\frac{1}{n}\mathbf{d}'\Sigma\mathbf{d}$, where

$$\mathbf{d} = (d_1, d_2, d_3, d_4)' = \left(\left. \frac{\partial g(\mathbf{m})}{\partial m_1} \right|_{\mathbf{m}=\Theta}, \dots, \left. \frac{\partial g(\mathbf{m})}{\partial m_4} \right|_{\mathbf{m}=\Theta} \right)'.$$

The elements of d are

$$d_1 = -d_2 = \frac{(\mu_y - \mu_x)}{\sigma_y \sigma_x (1 - \chi_a)^2},$$

$$d_3 = \frac{1}{2\sigma_y^2 \chi_a (1 - \chi_a)^2} - \frac{1}{2\sigma_y \sigma_x (1 - \chi_a)^2},$$

and

$$d_4 = \frac{1}{2\sigma_x^2 \chi_a (1 - \chi_a)^2} - \frac{1}{2\sigma_y \sigma_x (1 - \chi_a)^2}.$$

It can be shown that the variance of L is

$$\begin{aligned} \sigma_L^2 &= \frac{1}{n} d' \Sigma d \\ &= \frac{1}{n(1 - \chi_a)^2} \left[\chi_a^2 v^2 \left(\varpi + \frac{1}{\varpi} - 2\rho \right) + \frac{1}{2} \chi_a^2 \left(\varpi^2 + \frac{1}{\varpi^2} + 2\rho^2 \right) \right. \\ &\quad \left. + (1 + \rho^2) (\chi_a v^2 - 1) \right]. \end{aligned} \quad (2.49)$$

2.10 Proofs of Asymptotical Normality When Target Values Are Fixed

None of the proofs of asymptotic normality of the estimates of agreement indices other than CP with fixed target values have been given in the literature. The proofs here are much simplified compared to those in Section 2.9, because we are dealing with moments associated with Y only.

2.10.1 CCC and Precision Estimates

The $Z_{|X}$ transformation of the CCC estimate can be expressed as $Z_{|X} = g(\mathbf{m}_{|X})$, where

$$\mathbf{m}_{|X} = (m_{|X,1}, m_{|X,2}, m_{|X,3})' = (\bar{Y}, b_1, s_e^2)' \quad (2.50)$$

and

$$Z_{|X} = g(\mathbf{m}_{|X}) = \frac{1}{2} \ln \left[\frac{m_{|X,3} + m_{|X,2}^2 s_x^2 + s_x^2 + (m_{|X,1} - \bar{X})^2 + 2m_{|X,2} s_x^2}{m_{|X,3} + m_{|X,2}^2 s_x^2 + s_x^2 + (m_{|X,1} - \bar{X})^2 - 2m_{|X,2} s_x^2} \right].$$

Here $Z_{|X}$ is asymptotically normal with mean $\frac{1}{2} \ln \frac{1+\rho_{c|X}}{1-\rho_{c|X}}$ and variance $\frac{1}{n} \mathbf{d}'_{|X} \Sigma_{|X} \mathbf{d}_{|X}$, where

$$\Sigma_{|X} = \begin{pmatrix} \sigma_y^2(1-\rho^2) & 0 & 0 \\ 0 & \sigma_y^2(1-\rho^2)/s_x^2 & 0 \\ 0 & 0 & 2\sigma_y^4(1-\rho^2)^2 \end{pmatrix}. \quad (2.51)$$

The elements of $\mathbf{d}_{|X}$ are

$$d_{|X,1} = \frac{\rho_c^2(\mu_y - \bar{X})}{(1-\rho_c^2)^2 \rho \sigma_y s_x},$$

$$d_{|X,2} = -\frac{\rho_c}{(1-\rho_c^2)^2} \left(1 - \frac{1}{\rho_c \rho v}\right),$$

and

$$d_{|X,3} = \frac{\rho_c^2}{2(1-\rho_c^2)^2 \rho \sigma_y s_x}.$$

It can be shown that the variance of $Z_{|X}$ is

$$\sigma_{Z_{|X}}^2 = \frac{\rho_c^2(1-\rho^2)}{n(1-\rho_c^2)^2 \rho^2} \left[\varpi v^2 \rho_c^2 + (1-\rho_c \rho \varpi)^2 + \frac{\varpi^2 \rho_c^2(1-\rho^2)}{2} \right]. \quad (2.52)$$

When $\rho_c = \rho$, $v = 0$, and $\varpi = 1$, (2.52) degenerates to $\frac{1}{n} \left(1 - \frac{\rho^2}{2}\right)$, which is the variance of the Z transformation of the precision estimate.

2.10.2 MSD Estimate

From (2.50) we can write the natural logarithm of the MSD estimate, or $W_{|X} = \ln(e_{|X}^2)$, as

$$\begin{aligned} W_{|X} &= g(m_{|X,1}, m_{|X,2}, m_{|X,3}) \\ &= \ln \left[(m_{|X,1} - \bar{X})^2 + m_{|X,3} + s_x^2(1 - m_{|X,2})^2 \right]. \end{aligned}$$

By the delta method, $W_{|X}$ is asymptotically normal with mean $\ln(\varepsilon_{|X}^2)$ and variance $\frac{1}{n} \mathbf{d}'_{|X} \Sigma_{|X} \mathbf{d}_{|X}$, where $\Sigma_{|X}$ was shown in (2.51), and

$$\mathbf{d}_{|X} = \left(\frac{2(\mu_y - \bar{X})}{\varepsilon_{|X}^2}, \frac{-2(1-\beta_1)s_x^2}{\varepsilon_{|X}^2}, \frac{1}{\varepsilon_{|X}^2} \right)'.$$

It can be shown that the variance of $W_{|X}$ is

$$\sigma_{W_{|X}}^2 = \frac{2}{n} \left[1 - \frac{(\varepsilon_{|X}^2 - \sigma_e^2)^2}{\varepsilon_{|X}^4} \right].$$

2.10.3 CP Estimate

The proof can be seen in Lin, Hedayat, Sinha, and Yang (2002). Recall that in the regression model when target values are fixed, we assumed that e_Y has a normal distribution with mean 0 and variance σ_e^2 . Under this setup, the coverage probability of the i th observation is

$$\begin{aligned} \pi_{\delta_{0i}} &= \Pr(|Y_i - X_i| < \delta_0) \\ &= \Phi \left[\frac{\delta_0 - \beta_0 - (\beta_1 - 1)x_i}{\sigma_e} \right] - \Phi \left[\frac{-\delta_0 - \beta_0 - (\beta_1 - 1)x_i}{\sigma_e} \right]. \end{aligned}$$

We define the overall coverage probability as

$$\pi_{\delta_0|X} = \frac{1}{n} \sum_{i=1}^n \pi_{\delta_{0i}}. \quad (2.53)$$

Suppose that we have a random sample $\{(Y_i, X_i) \mid i = (1, \dots, n)\}$ and that β_0 , β_1 , and σ_e^2 are estimated by b_0 , b_1 , and s_e^2 . Then b_0 and b_1 are independent of s_e . An estimate of $\pi_{\delta_{0i}}$ is

$$p_{\delta_{0i}} = \Phi \left[\frac{\delta_0 - b_0 - (b_1 - 1)x_i}{s_e} \right] - \Phi \left[\frac{-\delta_0 - b_0 - (b_1 - 1)x_i}{s_e} \right],$$

and an estimate of $\pi_{\delta_0|X}$ is

$$p_{\delta_0|X} = \frac{1}{n} \sum_{i=1}^n p_{\delta_{0i}}.$$

By the same method as shown in Section 2.9.3, it can be shown that

$$E(p_{\delta_0|X}) = \pi_{\delta_0|X} + O\left(\frac{1}{n}\right),$$

and that the asymptotic variance of $p_{\delta_0|X}$ is

$$\sigma_{p_{\delta_0|X}}^2 = \frac{1}{n} \left[\frac{C_0^2}{n^2} + \frac{(C_0 \bar{X} - C_1)^2}{n^2 s_x^2} + \frac{C_2^2}{2n^2} \right] + O\left(\frac{1}{n^2}\right),$$

where C_0 , C_1 and C_2 are defined in (2.19), (2.20), and (2.21).

2.10.4 Accuracy Estimate

Here we use the same $\Sigma_{|X}$ as in (2.51). The elements of $\mathbf{d}_{|X} = (d_{|X1}, d_{|X2}, d_{|X3})'$ become

$$d_{|X1} = -\frac{(\mu_y - \bar{X})^2}{\sigma_y s_x (1 - \chi_a)^2},$$

$$d_{|X2} = -\frac{s_x \beta_1}{\sigma_y (1 - \chi_a)^2} + \frac{s_x^2 \beta_1}{\sigma_y^2 \chi_a (1 - \chi_a)^2},$$

and

$$d_{|X3} = -\frac{1}{2\sigma_y s_x (1 - \chi_a)} + \frac{1}{2\sigma_y^2 \chi_a (1 - \chi_a^2)}.$$

Therefore, the variance of $L_{|X}$ is

$$\sigma_{L_{|X}}^2 = \frac{v^2 \varpi \chi_a^2 (1 - \rho^2) + \frac{1}{2} (1 - \varpi \chi_a)^2 (1 - \rho^4)}{n(1 - \chi_a)^2}.$$

2.11 Other Estimations and Statistical Inference Approaches

Estimations of agreement indices presented in this chapter are based on moment estimations by replacing proposed indices with their respective sample counterparts. Statistical inferences based on these estimates are carried out by the routine delta method.

King and Chinchilli (2001a) proposed estimations and statistical inferences for the CCC based on the U-statistic outlined by Davis and Quade (1968). Barnhart and Williamson (2001) proposed estimations and statistical inferences for the CCC based on the GEE methodology outlined by Liang and Zeger (1986). Barnhart, Haber, and Song (2002) later extended the GEE methodology for multiple assays or raters. Both U-statistic and GEE methodologies have the advantage of addressing both estimation and statistical inference simultaneously based on established general formulas.

Carrasco and Jover (2005) proposed to use the maximum likelihood (ML) method with a mixed effect model. All of the above approaches were proposed for CCC only, and have been extended to be applicable when we have multiple assays or raters. However, none of the above has addressed cases in which the target values are fixed. The U-statistic and GEE methodology for the CCC are also valid when we have categorical data, to be discussed in Chapters 3 and 5.

2.11.1 *U-Statistic for CCC*

Suppose we have $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ random samples from n samples or subjects. For $i, j = 1, 2, \dots, n$, let

$$\varphi_{1ij} = (X_i - Y_i)^2 + (X_j - Y_j)^2 - (X_i + Y_i)^2 - (X_j + Y_j)^2,$$

$$\varphi_{2ij} = X_i^2 + X_j^2 + Y_i^2 + Y_j^2,$$

$$\varphi_{3ij} = (X_i - Y_j)^2 - (X_i + Y_j)^2 + (X_j - Y_i)^2 - (X_j + Y_i)^2,$$

$$U_1 = \frac{\sum_{ij} \varphi_{1ij}}{2n(n-1)},$$

$$U_2 = \frac{\sum_{ij} \varphi_{2ij}}{n(n-1)},$$

and

$$U_3 = \frac{\sum_{ij} \varphi_{3ij}}{2n(n-1)}.$$

King and Chinchilli (2001a) showed that the CCC estimate can be written in terms of functions of the U-statistics as

$$r_c = \frac{H}{G} = \frac{(n-1)(U_3 - U_1)}{U_1 + nU_2 + (n-1)U_3}. \quad (2.54)$$

They further showed that the Z -transformation of r_c by the delta method has asymptotic normal distribution with mean $\frac{1}{2} \tanh^{-1}(\rho_c)$, and variance

$$\sigma_Z^2 = \frac{\rho_c^2}{(1 - \rho_c^2)^2} \left[\frac{\sigma_H^2}{H^2} - \frac{2\sigma_{HG}}{HG} + \frac{\sigma_G^2}{G^2} \right], \quad (2.55)$$

where

$$\sigma_H^2 = (n-1)^2 [V(U_3) + V(U_1) - 2\text{cov}(U_3, U_1)],$$

$$\begin{aligned} \sigma_G^2 = & (n-1)^2 V(U_3) + V(U_1) + n^2 V(U_2) + 2(n-1)\text{cov}(U_3, U_1) \\ & + 2n(n-1)\text{cov}(U_3, U_2) + 2n\text{cov}(U_3, U_2), \end{aligned}$$

and

$$\begin{aligned} \sigma_{HG} = & -(n-1)(n-2)\text{cov}(U_3, U_1) + n(n-1)\text{cov}(U_3, U_2) + (n-1)^2 V(U_3) \\ & -(n-1)V(U_1) - n(n-1)\text{cov}(U_2, U_1). \end{aligned}$$

The variance-covariance matrix of $\mathbf{U} = (U_1, U_2, U_3)'$, denoted by V , can be obtained as follows. Let

$$\boldsymbol{\varphi}_{\bullet i} = (\varphi_{1i}, \varphi_{2i}, \varphi_{3i})',$$

where

$$\varphi_{1i} = \frac{\sum_j \varphi_{1ij}}{(n-1)}, \quad \varphi_{2i} = \frac{\sum_j \varphi_{2ij}}{(n-1)}, \quad \varphi_{3i} = \frac{\sum_j \varphi_{3ij}}{(n-1)}.$$

Then we have

$$V = \frac{4}{n^2} \sum_i (\varphi_{*i} - U)' (\varphi_{*i} - U).$$

2.11.2 GEE for CCC

Barnhart and Williamson (2001) first proposed to use GEE for statistical estimation and inference for CCC. They used three sets of GEE equations, one for estimating means accounting for covariates, one for estimating variances without accounting for covariates, and one for estimating the Z -transformed CCC. Variance-covariance matrices of the above estimates can be estimated, and the delta method can be applied to obtain the asymptotic normality of the Z -transformed CCC estimate.

Suppose we have $(Y_{11}, Y_{21}), (Y_{12}, Y_{22}), \dots, (Y_{1n}, Y_{2n})$ random samples from n samples or subjects. Let \mathbf{Y}_i be the 2×1 vector that contains the two readings and let the $2 \times p$ matrix \mathbf{X}_i be the corresponding p covariates for the i th sample, where the first column of \mathbf{X}_i is a vector of all ones representing an intercept term. Let $\mathbf{Y}_i = (Y_{1i}, Y_{2i})'$ and $\boldsymbol{\beta}$ be a 2×1 marginal parameter vector. The three GEE equations are shown below.

In the first set of equations, the marginal mean vector of \mathbf{Y}_i is $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$, and the parameter estimates of $\boldsymbol{\beta}$ are obtained by

$$\sum_1^n \mathbf{D}_i' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = 0, \quad (2.56)$$

where $\mathbf{D}_i = \frac{d\boldsymbol{\mu}_i}{d\boldsymbol{\beta}}$ and \mathbf{V}_i is the working covariance matrix for \mathbf{Y}_i (Zeger and Liang 1986).

Let σ_1^2 and σ_2^2 be variances of \mathbf{Y}_1 and \mathbf{Y}_2 . In the second set of equations, the variances of \mathbf{Y}_1 and \mathbf{Y}_2 without accounting for covariates are estimated by

$$\sum_1^n \mathbf{F}_i' \mathbf{H}_i^{-1} (\mathbf{Y}_i^2 - \sigma_i^2(\sigma^2, \boldsymbol{\beta})) = 0, \quad (2.57)$$

where $\mathbf{F}_i = \frac{d\sigma_i^2}{d\sigma^2}$ and \mathbf{H}_i is the working covariance matrix for \mathbf{Y}_i^2 , $\sigma_i^2 = \sigma^2 + \boldsymbol{\mu}_i^2$, and $\sigma^2 = (\sigma_1^2, \sigma_2^2)'$. In solving these equations, the diagonal components of \mathbf{H}_i are assumed normal even if \mathbf{Y}_i is not normally distributed.

Let $\theta_i = E(Y_{1i} Y_{2i})'$ and let Z be the Z -transformed CCC. In the third equation, Z is estimated by

$$\sum_1^n C_i W_i^{-1} (Y_1 Y_2 - \theta_i(Z, \beta, \sigma^2)) = 0, \quad (2.58)$$

where $C_i = \frac{d\theta_i}{dZ}$ and W_i is the variance of θ_i .

This GEE method and the U-statistic method yield the same CCC estimate as proposed by Lin (1989), but the variances of the CCC estimate are slightly different because these two methods do not assume normality, while the method by Lin assumes normality in the computation of the variance of the CCC estimate.

2.11.3 Mixed Effect Model for CCC

Carrasco and Jover (2005) proposed to use the maximum likelihood (ML) or restricted ML (RML) method through a mixed effect model. Robieson (1999) and Carrasco and Jover (2005) showed that the CCC is a special form of ICC under the mixed effect model of random subject effect with the variance σ_α^2 , residual effect with the variance σ_e^2 , and fixed assay or rater effect with the mean square σ_β^2 , when σ_β^2 is included in the denominator. Specifically, the CCC can be expressed as

$$\rho_c = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_e^2 + \sigma_\beta^2}. \quad (2.59)$$

In Section 3.1.3, we will revisit this coefficient in detail. Carrasco and Jover (2005) proceeded to use the delta method for statistical inference after estimating the variance–covariance matrix of the variance components through ML or RML.

This method does not yield the same CCC estimate as proposed by Lin (1989), and the variance of the CCC estimate can be slightly different, because this method assumes normality through MLE or RMLE.

2.11.4 Other Methods for TDI and CP

Compared to CCC, there have been fewer contributions related to TDI and CP. Some of the methods related to TDI and CP in the literature are pointed out in the last paragraph of Section 2.13. Most of those articles use more complicated iterative methods to fine-tune TDI and CP as well as their confidence intervals.

2.12 Discussion

2.12.1 Absolute Indices

Three agreement statistics, MSD, TDI, and CP, are unscaled indices, which do not depend on the between-subject variation. TDI and CP attempt to capture a large proportion (CP) of observations that are within a certain deviation (TDI) from their target values. We can compute CP for a given TDI, denoted by CP_{δ_0} , or compute TDI for a given CP, denoted by TDI_{π_0} . When the error structure is proportional, we apply a log transformation to the data, and the resulting TDI is then antilog transformed. When we subtract 1.00 from this antilog-transformed value and multiply by 100, it becomes a percent change ($TDI\%_{\pi_0}$) rather than an absolute difference (Lin 2000, 2003, Lin, Hedayat, Sinha, and Yang 2002). This means that $100\pi_0\%$ of observations are within $TDI\%_{\pi_0}$ of the target values. TDI and CP offer the most intuitively clear interpretation and have better power for statistical inference, yet they do not have precision and accuracy components. Also, Lin, Hedayat, Sinha, and Yang (2002) and Lin, Hedayat, and Wu (2007) used approximations and assumed normality to perform estimations and statistical inferences for TDI and CP. When the data are not normally or log-normally distributed, a transformation to bring the data closer to normality might be necessary.

TDI and CP are mirrored statistics. The former requires a given coverage probability to compute the absolute difference or percent change. The latter requires a given absolute difference or percent change to compute the coverage probability. The former has the advantage for the following reason:

- TDI can discriminate among assays with much better agreement than CP, because in these cases, CP values are near one.
- When there is no hard allowance available, one can still compute TDI at $CP = 0.8$ or 0.9 , but one cannot compute CP without a reasonably given TDI value.

2.12.2 Relative Indices Scaled to the Data Range

Due to its equivalence to kappa and weighted kappa as well as its close tie to ICC, the CCC is perhaps the most popular index among statisticians for assessing agreement. The CCC and precision and accuracy coefficients are ICC-like (Lin, Hedayat, and Wu 2007), and are scaled (relative) to the total variation, especially the between-sample variation. This property is appealing when we wish to assess agreement over the entire reasonable value range from normal to abnormal. Comparisons among any of these three agreement statistics are valid only with similar study ranges, which is proportional to the between-sample variation. When the study range is fixed and meaningful, the CCC and precision and accuracy coefficients offer meaningful geometric interpretations. It is important to report the study range when reporting these statistics. Good agreement over a small range

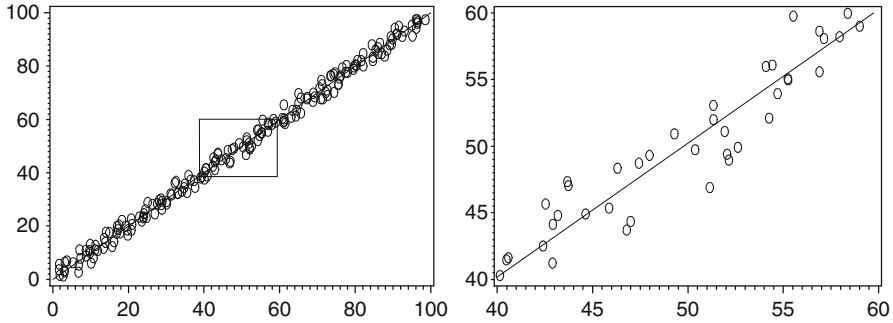


Fig. 2.13 Agreement over larger and shorter analytical ranges. (a) Larger analytical range. (b) Shorter analytical range

of measurements cannot be extrapolated to conclude good agreement over a larger range of measurements.

In other words, we should not conduct a method-comparison experiment over a range similar to the range of intrasample random fluctuation (Lin and Chinchilli 1997). As an illustrative example, Fig. 2.13a shows an artificial result of a good agreement over a desirable analytical range. If we study a subset of the data, limited to a much shorter analytical range (the box portion of Fig. 2.13a), which is magnified in Fig. 2.13b, any correlation coefficient would be much smaller.

2.12.3 Variances of Index Estimates Under Random and Fixed Target Values

Apart from CP, all of the agreement coefficients defined in this chapter have the same coefficient estimates regardless whether the target values are assumed random or fixed. The CP estimates under random and fixed target values are asymptotically the same. The variance of each of these coefficient estimates under the random target assumption is always larger than the variance under the fixed target assumption. These are evident by comparing (2.3) and (2.4) for the log MSD (TDI), (2.13) and (2.14) for the logit CP, (2.29) and (2.30) for the Z-transformed CCC, (2.31) and (2.32) for the logit accuracy coefficient, and the formulas in the text after (2.32) for the Z-transformed precision coefficient. Therefore, the confidence limits of an agreement coefficient would be closer to its coefficient estimate under the fixed target assumption than under the random target assumption.

2.12.4 Repeated Measures CCC

There are two types of repeated-measures data for agreement assessment that we often encounter. For one type, the between-sample variation forms the data range.

For the other, the repeated measures form the data range. An example of the first case is to have each subject's blood pressures measured over time, which is what is usually encountered in practice. Many tools are available for this type of repeated measure, which can be found in the Section 2.13. An example of the second case is to have each sample taken from a homogeneous population and to perform serial dilutions that form the data range. Such serial dilutions are uniform across all homogeneous samples. In this case, we could compute agreement coefficient estimates for each sample, and treat these estimates as random samples from a population. We then compute means and confidence limits based on the respective transformations of the agreement statistics. Antitransformation of these limits would be their respective confident limits. Such an approach is valid if we don't have missing data. We may also follow the more detailed approach proposed by Chinchilli, Martel, Kumanyika, and Lloyd (1996).

2.12.5 Data Transformations

The CCC and accuracy and precision coefficient estimates are in general quite robust against moderate deviation from normality. If not, there are tools based on robust estimates by King and Chinchilli (2001b) and based on a nonparametric approach by Guo and Manatunga (2007).

The TDI and CP estimates are heavily dependent on the normality or log-normality assumption. When there is evidence that data are not normally distributed for the constant random error case, and not log-normally distributed for the proportional random error case, data transformation might be necessary for the robustness of TDI and CP estimates. In this case, see Lin and Vonesh (1989) for the transformation approach that minimizes the MSD between the ordered observed and theoretical quantiles.

2.12.6 Missing Data

In this book we deal only with no-missing-data cases. In this chapter and Chapter 3, we discuss the case of paired assays or raters, and we often do not encounter a large amount of missing data in practice. Therefore, deleting cases with missing data is a reasonable approach. Missing data situation can sometimes be an issue in practice as pointed out in Chapters 5 and 6 when we have multiple raters with replicates. Research in the social and psychological sciences and in clinical trials may often encounter missing data. Approaches that can handle missing data should be an interesting area of research.

2.12.7 Account for Covariants

Covariate adjustment has been a controversial topic in assessing agreement. Proponents argue that without such adjustment, CCC and thus accuracy and precision coefficients are artificially inflated. Opponents argue that such adjustment artificially decreases the CCC because the data range is being reduced, as seen in an earlier paragraph about the effect of the data range. Opponents further argue that assay or rater agreement that could depend on covariants cannot be judged as having good agreement. In addition, often covariants are selected to cover a desirable data range, as in the use a variety of species in Example 2.8.5. In this case, the covariate adjustment related to species can be misleading. We believe that there are cases in which covariate adjustment is meaningful. However, we recommend that the practice of covariate adjustment be used with caution.

The GEE methodology proposed by Barnhart and Williamson (2001) allows for covariate adjustment, but only for means, not for variances and covariances. To adjust for variances and covariances, a simple and reasonable way is to perform the linear regression for each assay or rater, then use the intercept or overall mean estimate plus the residual of each assay or rater as the adjusted dependent variable. The adjusted dependent variable of each assay or rater would still contain the subject-to-subject variation without the effect of covariates. We can then perform the estimations and statistical inferences of the agreement indices based on these adjusted dependent variables. Such an approach is appropriate when covariates are subject-specific, such as age and gender. This approach is not applicable when we have fixed target values. However, we have rarely encountered covariates for which the target values are fixed.

2.13 A Brief Tour of Related Publications

Chapter 2 is based largely on the materials in Lin, Hedayat, Sinha, and Yang (2002). For an earlier introduction of CCC and precision and accuracy coefficients, see Lin (1989), and for TDI, see Lin (2000).

The method of Bland and Altman (1986) for assessing agreement uses a meaningful graphical approach and computes the confidence limits from the paired differences. Because of its simplicity, this approach has been quite popular among medical researchers. This approach lacks a specific index to summarize the degree of agreement, and thus statistical inferences about the estimate cannot be performed. Bland and Altman later (1999) improved on their approach with statistical inference. Their approaches are similar to our TDI. The major difference between our TDI and their approaches is that we capture a majority of observations from their respective individual target values, while their approaches capture the paired differences from the mean of paired differences.

Chinchilli, Martel, Kumanyika, and Lloyd (1996) addressed repeated-measures CCC, which is the weighted average of CCCs across subjects. Vonesh, Chinchilli, and Pu (1996) and Vonesh and Chinchilli (1997) modified CCC for goodness-of-fit. King and Chinchilli (2001a) used a U-statistics framework for CCC, which includes the generalization of CCC for multiple assays or raters, and the approach is applicable to categorical data. Barnhart and Williamson (2001) used GEE to estimate CCC, which also can be extended to include the generalization of CCC for multiple assays or raters (Barnhart, Haber, and Song 2002). In our opinion, these GEE methodologies are also applicable for categorical data, although the authors did not make such a claim.

King and Chinchilli (2001b) proposed a robust estimation of CCC through an absolute loss function or M-estimation by U-statistics. Li and Chow (2005) used weighted CCC by kernel density for repeated-measures image data. Quiroz (2005) proposed to assess agreement using the CCC in a repeated-measurement design. Liu, Du, Teresi, and Hasin (2005) proposed CCC for survival data. Barnhart, Song, and Lyles (2005) proposed assay validation for left-censored data. Carrasco and Jover (2005) proposed to use the maximum likelihood (ML) or restricted ML (RML) method through a mixed-effect model, which is applicable to multiple assays or raters. Robieson (1999) and Carrasco and Jover (2005) showed that the CCC is a special form of ICC under the mixed-effect model of random subject effect and residual effect and fixed assay or rater effect when the fixed assay or rater effect is included in the denominator of the ICC. King, Chinchilli, and Carrasco (2007) proposed another approach of repeated-measures CCC. They used the population estimates, rather than subject-specific estimates proposed earlier by Chinchilli, Martel, Kumanyika, and Lloyd (1996), to construct a repeated-measures CCC. King, Chinchilli, Wang, and Carrasco (2007) presented a class of repeated-measures CCC. Guo and Manatunga (2007) proposed using nonparametric estimation of the CCC under univariate censoring. Carrasco, Luis, King, and Chinchilli (2007) compared concordance correlation coefficient estimating approaches with skewed data. Williamson, Crawford, and Lin (2007) presented a permutation testing for comparing dependent CCCs. Quiroz and Burdick (2009) again proposed an assessment of individual agreement with repeated measurements based on generalized confidence intervals. Carrasco, King, and Chinchilli (2009) again proposed repeated-measures CCC estimated by variance components. Hiriot and Chinchilli (2010) proposed matrix-based concordance correlation coefficient for repeated measures. Helenowski, Vonesh, Demirtas, Rademaker, Ananthanarayanan, Gann, and Jovanovic (2011) extended CCC by allowing for different spatial variance-covariance structures of the data. They proposed a general concordance correlation matrix representing pairwise CCCs along with an overall CCC.

There have been relatively fewer publications for TDI and CP, which are summarized below. Wang and Gene Hwang (2001) proposed a nearly unbiased test (NUT) based on CP for the application of individual bioequivalence. Choudhary and Nagaraja (2007) proposed an exact test and modified NUT for CP and TDI for data with a small sample size (<30) that need to be solved numerically through iterations, and a bootstrap estimation for data with a moderate sample size. Hedayat, Lou, and

Sinha (2009) introduced CP involving multiple assays or raters. Escaramis, Ascaso, and Carrasco (2010) simplified the approach by Choudhary and Nagaraja (2007) for TDI using a tolerance limit approach through iterations. Choudhary (2008) proposed a tolerance interval approach for assessment of agreement in method comparison studies with repeated measurements. Choudhary (2008) proposed the tolerance approach with left censored data.

Statistical Tools for Measuring Agreement

Lin, L.; Hedayat, A.S.; Wu, W.

2012, XVI, 161 p., Hardcover

ISBN: 978-1-4614-0561-0